DEC 11 1941
DUKE ADMIN LIBRARY

Journal of the
AMERICAN STATISTICAL ASSOCIATION

CONTENTS *for* DECEMBER 1941

Birth Rates and the Interwar Business Cycles *Virginia L. Galbraith and Dorothy S. Thomas* 465

A Sequence of Historical Random Events: Do Jesuits Die in Three's? *J. Solterer* 477

Relation Between Seasonal Amplitudes and the Level of Production—An Application to the Production of Steel Ingots *Gustave S. Juliber* 485

Probabilities in Logarithmic Skewed Distributions *G. R. Davies and R. H. Smith* 493

Linear Regression When the Standard Deviations of Arrays Are Not All Equal *G. A. Baker* 500

Accumulating Cubes with Punch Cards *Marion M. Sandomire* 507

Fitting the Exponential Function and the Gomperts Function by the Method of Least Squares *Paul Matthew Stoner* 515

Further Notes on the Difference Between Index-Number Formulas *Irving H. Siegel* 519

The Use of Grouped Measurements *Howard L. Jones* 525

A New Method for Obtaining the Inverse Matrix *M. D. Bingham* 530

A Punch Card Designed to Contain Written Data and Coding *Joseph Berkson* 535

Notes and Discussions 539

Book Reviews 548

VOLUME 36 NUMBER 216

American Statistical Association

Organized November 27, 1839
Incorporated 1841

The American Statistical Association is a scientific and educational organization. Its membership is not confined to professional statisticians but includes economists, business executives, research directors, government officials, university professors, and other persons who are seriously interested in the application of statistical methods to practical problems, in the development of more useful methods, and in the improvement of basic statistical data. Engineers, mathematicians, biologists, actuaries, sociologists, psychologists, and representatives of many other professions are included in the membership of the Association. Information about the Association and membership application forms may be secured from the Secretary, R. L. Funkhouser, 1626 K Street, N.W., Washington, D. C.

The 103rd Annual Meeting of the American Statistical Association will be held in New York City, December 27 to 30, 1941. With a few exceptions the sessions are open to the public.

Chapters of the Association have been organized in important cities in the United States. Information concerning the meetings of these chapters may be secured from the District Representatives listed on another page.

The Editors welcome the submission of articles and notes for possible publication in the JOURNAL. Manuscripts should be sent to the Editor, Journal of the American Statistical Association, 1626 K Street, N.W., Washington, D. C. Books for review should be sent to the Review Editor, Glenn E. McLaughlin, 210 Normandy Building, 1626 K Street, N.W., Washington, D. C. Authors who wish suggestions about the preparation of manuscripts and charts or information about editorial policies should address their inquiries to the Editor.

Subscription rate, $6.00 per annum. Single numbers of volumes 23 to 35 $1.50 per copy postpaid, except Proceedings Supplements to volumes 23 to 30, $2.00 per copy postpaid. Prices for sets and discounts for orders in quantity available on request.

Published Quarterly by the AMERICAN STATISTICAL ASSOCIATION

Publication Office: 450 Ahnaip Street, Menasha, Wis. Editorial Office: 1626 K Street, N.W., Washington, D.C. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in paragraph 4, section 538, P. L. & R., authorized March 25, 1936. Entered as second class matter at the post office at Menasha, Wisconsin.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

DECEMBER, 1941 · $1.50 PER COPY · $6.00 PER ANNUM · VOL. 36 · NO. 216

CONTENTS

ARTICLES

Birth Rates and the Interwar Business Cycles . . . . . VIRGINIA L. GALBRAITH AND DOROTHY S. THOMAS 465

A Sequence of Historical Random Events: Do Jesuits Die in Three's? . . . . . J. SOLTERER 477

Relation Between Seasonal Amplitudes and the Level of Production—An Application to the Production of Steel Ingots . GUSTAVE S. JULIBER 485

Probabilities in Logarithmic Skewed Distributions . . . . . G. R. DAVIES AND R. H. SMITH 493

Linear Regression When the Standard Deviations of Arrays Are Not All Equal . . . . . G. A. BAKER 50$_0$

Accumulating Cubes with Punch Cards . . . MARION M. SANDOMIRE 507

Fitting the Exponential Function and Gompertz Function by the Method of Least Squares . . . . . PAUL MATTHEW STONER 515

Further Notes on the Difference Between Index-Number Formulas . . . . . IRVING H. SIEGEL 519

The Use of Grouped Measurements . . . . . HOWARD L. JONES 525

A New Method for Obtaining the Inverse Matrix . . M. D. BINGHAM 530

A Punch Card Designed to Contain Written Data and Coding . . . . . JOSEPH BERKSON 535

NOTES AND DISCUSSIONS

Comments on Dr. Madow's "Note on Tests of Departure from Normality" with Some Remarks Concerning Tests of Significance
I . . . . . JOSEPH BERKSON 539
II. . . . . R. C. GEARY 541

Roland Post Falkner, 1866–1940 . . . . . F. LESLIE HAYFORD 543

Sumner B. Pearmain, 1859–1941 . . . . . ROSWELL F. PHELPS 545

Josiah Charles Stamp, 1880–1941 . . . . . WALTER F. WILLCOX 546

Gusztav Thirring, 1861–1941 . . . . . FRANK A. ROSS 547

BOOK REVIEWS

ALBERT, A. ADRIAN. *Introduction to Algebraic Theories.* Franklin E. Satterthwaite 576

ARNESON, G., see James, C. L.

BARNETT, GEORGE E., see Evans, George H., Jr.

BARNETT, PAUL. *Business-Cycle Theory in the United States, 1860–1900.* Elmer C. Bratt 572

BECKETT, GRACE. *The Reciprocal Trade Agreements Program.* E. Dana Durand 548

BERTRAND, DANIEL; EVANS, W. DUANE, and BLANCHARD, E. L. *The Motion Picture Industry.* Evan B. Alderfer 565

BLAIR, JOHN M., and REESIDE, ARTHUR. *Price Discrimination in Steel.* Marion W. Worthing 566

BLANCHARD, E. L., see Bertrand, Daniel

CORNELL, FRANCIS G., see Mort, Paul R.

DAVENPORT, DONALD H., and GESELL, GERHARD A. *Families and Their Life Insurance.* J. Edward Hedges 564

DICKENS, PAUL D., see Gilbert, Milton.

DOANE, ROBERT R. *The Anatomy of American Wealth.* Maxine Y. Sweezy 569

Continued on next page

BOOK REVIEWS (*Continued*)

ESTEY, JAMES A. *Business Cycles: Their Nature, Cause, and Control.* Paul Barnett 571

EVANS, GEORGE H., JR., and BARNETT, GEORGE E. *Principles of Investment.* Melchior Palyi 556

EVANS, W. DUANE, see Bertrand, Daniel.

FOOD RESEARCH INSTITUTE, STANFORD UNIVERSITY. *Wheat Studies of the Food Research Institute,* Vol. XVII, No. 7. N. Jasny 554

GESELL, GERHARD A., and HOWE, ERNEST J. *Study of Legal Reserve Life Insurance Companies.* J. Edward Hedges 564

GESELL, GERHARD A., see Davenport, Donald H.

GILBERT, MILTON, and DICKENS, PAUL D. *Export Prices and Export Cartels.* Charles A. Welsh 561

HAVEN, T. KENNETH. *Investment Banking Under the Securities and Exchange Commission.* Raymond W. Goldsmith 577

HOFFMAN, A. C. *Large-Scale Organization in the Food Industries.* Evan B. Alderfer 565

HOWE, ERNEST H., see Gesell, Gerhard A.

HUNTINGTON, EDWARD V. *Methods of Apportionment in Congress.* Churchill Eisenhart 568

JAMES, C. L., WELSH, E. C., and ARNESON, G. *Industrial Concentration and Tariffs.* Charles A. Welsh 561

KREPS, THEODORE J. *Measurement of the Social Performance of Business.* Spurgeon Bell 559

MERWIN, CHARLES L., Jr. *Financial Characteristics of American Manufacturing Corporations.* Charles A. Bliss. 563

MILLET, JOHN I. *Bank Audits and Examinations.* Kenneth L. Trefftzs . 558

MORT, PAUL R., and CORNELL, FRANCIS G. *American Schools in Transition.* Douglas E. Scates 573

NATIONAL RESOURCES COMMITTEE. *Urban Planning and Land Policies.* Homer Hoyt 552

PEACH, WILLIAM NELSON. *The Security Affiliates of National Banks.* Kenneth L. Trefftzs 557

PFIFFNER, JOHN M. *Research Methods in Public Administration.* George A. Graham 549

REESIDE, ARTHUR, see Blair, John M.

SECURITIES AND EXCHANGE COMMISSION. *Statistics of American Listed Corporations.* William A. Spurr 578

SIMON, LESLIE E. *An Engineers' Manual of Statistical Methods.* Eugene L. Grant 574

SLICHTER, SUMNER H. *Economic Factors Affecting Industrial Relations Policy in National Defense.* Francis M. Boddy 550

SPURRIER, LEO. *Common Stocks and Bonds as Long-Term Investments.* Alfred Cowles 555

STIGLER, GEORGE J. *Production and Distribution Theories, the Formative Period.* O. J. McDiarmid 570

TAITEL, MARTIN. *Profits, Productive Activities, and New Investment.* Solomon Fabricant 562

*TEMPORARY NATIONAL ECONOMIC COMMITTEE. *Statement on Life Insurance,* Monograph No. 28-A. J. Edward Hedges 564

TEMPORARY NATIONAL ECONOMIC COMMITTEE. *Reports of the Federal Trade Commission,* Monograph No. 36. Evan B. Alderfer 565

TEMPORARY NATIONAL ECONOMIC COMMITTEE. *The Basing Point Problem,* Monograph No. 42. Marion W. Worthing 566

THOMAS, DOROTHY SWAINE. *Social and Economic Aspects of Swedish Population Movements, 1750–1933.* W. Lewis Abbott 551

TURNER, H. HAINES. *Case Studies of Consumers' Cooperatives.* Carl N. Schmalz 553

WELSH, E. C., see James, C. L.

WEST, CARL J. *Introduction to Mathematical Statistics.* Franklin E. Satterthwaite 575

Publications Received 580

* For TNEC Monographs Nos. 2, 6, 7, 10, 12, 15, 28, 35, 41, and 43, see Davenport, Gilbert, Kreps, James, Taitel, Merwin, Gesell Hoffman, Blair, and Bertrand.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 216 DECEMBER, 1941 *Volume 36*

BIRTH RATES AND THE INTERWAR BUSINESS CYCLES

By Virginia L. Galbraith and Dorothy S. Thomas
University of California

As YULE remarked in 1906: "It seems a curious fact that while the relations of the marriage-rate to economic factors have comparatively frequently drawn the attention of statistical writers, their effects on the birth-rate do not appear to have been so fully discussed. Yet such a discussion must form the very basis of the theory of population in a modern industrial state."[1] Yule's own analysis of this relationship was one of the first which met the essential statistical requirements of eliminating secular trends and computing correlation coefficients between the deviations from the trends of the series under consideration. Analyzing first the relationship between indexes of business cycles, namely a price index (1865–1896) and an index of unemployment (1870–1895), and marriage rates for England and Wales, he found correlation coefficients of +.795 and −.873 respectively. These results confirmed the results of earlier studies by Hooker[2] and March[3] with similar data and similar methods. Proceeding then to an analysis of the correlation of the marriage rate and the birth rate, for the period 1850–1896, he found a maximum correlation of +.479 with the birth rate lagging two years, and concluded that the oscillations in the birth rate were influenced by economic oscillations not only indirectly through their response to oscillations in the marriage rate, but also directly. Because of defects in the data at his disposal, he confessed himself unable to make a clearcut analysis of the causal nexus. However, by calculating the course of marriage rate deviations and deducing the birth rate deviations for each year "using rough graphically interpolated values for the births per thousand married women,"[4] he esti-

[1] G. Udny Yule, "On the Changes in the Marriage- and Birth-Rates in England and Wales During the Past Half Century; with an Inquiry as to Their Probable Causes," *Journal of the Royal Statistical Society*, vol. 69, 1906, p. 122.

[2] R. H. Hooker, "Correlation of the Marriage-Rate with Trade," *Journal of the Royal Statistical Society*, vol. 64, 1901, pp. 485–492.

[3] Lucien March, "Comparaison numérique de courbes statistiques," *Journal de la Société de Statistique de Paris*, vol. 46, 1905, pp. 255–277; 306–311.

[4] Yule, *loc. cit.* (1906), p. 124.

mated that the direct factor possibly accounted for half of the amplitude of the oscillations. In a later paper (1925),[5] he did not carry this type of analysis further, but correlated marriage, birth (with a lag of two years) and death rate deviations with a foreign trade series for 1859–1908. The coefficients of correlation were +.78, +.47, and +.18 respectively. He concluded that, in the case of death rates, there was not "definite consonance between the waves. . . . The waves in the marriage-rate, on the other hand . . . show a remarkable consilience with those in trade, and the birth-rate also exhibits, up till about 1895 at least, waves with a lag of two years, or rather more, on the trade waves. . . . It suffices to note that here again it is the 'preventive check' rather than the 'positive check' which appears to be operative."[6]

One of the earliest analyses based on American data was made by Ogburn and Thomas in 1922.[7] Using marriage and birth rates for the period 1870–1920 for six states (Connecticut, Massachusetts, Michigan, New Hampshire, Rhode Island, and Vermont) and a composite index of business cycles for the United States, they found a correlation coefficient of +.87 between business cycles and marriage rates and of +.33 between business cycles and birth rates, when births lagged by one year. They, too, concluded that the latter relationship was partly direct, partly to be attributed to the "fact that marriages are so highly correlated with business conditions."[8] Thomas, in 1925,[9] correlated birth rates in England and Wales with a composite index of British business cycles for the period 1854–1913, after first demonstrating a correlation of +.67 between business cycles and marriage rates. The birth rate correlation amounted to only +.29 with two years' lag, and to +.30 with three years' lag behind business cycles. She, too, emphasized the "preventive check," giving considerable weight, in her interpretations, to the fact that the correlations between marriages and business cycles tended to diminish during the period under consideration, while those between births and business cycles rose steadily to a maximum of +.64 for the subperiod 1895–1913. "It suggests the interesting hypothesis that, whereas a business depression formerly led to 'prudential restraint' through abstention from marriage, it now leads to a more widespread and deliberate use of birth control." In common

[5] G. Udny Yule, "The Growth of Population and the Factors Which Control It," *Journal of the Royal Statistical Society* vol. 88, 1925, pp. 1–58.

[6] Yule, *loc. cit.* (1925), p. 30.

[7] William F. Ogburn and Dorothy S. Thomas, "The Influence of the Business Cycle on Certain Social Conditions," *Quarterly Publications of the American Statistical Association*, vol. 18, 1922, pp. 324–340.

[8] *Ibid.*, p. 339.

[9] Dorothy S. Thomas, *Social Aspects of the Business Cycle*, London, 1925.

with the other investigators mentioned, she assumed, without rigorous proof, that "the response of the birth-rate to the business cycle, then, is probably partly direct, through the differential conscious control of births in prosperity and depression, and partly indirect, through a reflection of the fluctuations in the marriage rate."[10]

Other studies have appeared sporadically in the course of the past two decades. In general they have confirmed the tendencies noted by the earlier investigators, but, with one exception, which will be noted in the following paragraph, have added nothing new. They were hampered by the fact that the World War had introduced a major break in the continuity of the time series and that not enough time had passed to give a new series that would be statistically significant. With the onset of a second World War, it is now apparent that the experience of the two decades separating the two wars will, whether long enough for statistical significance or not, make possible the only analysis for some time to come of the "normal" impact of business cycles upon birth rates. This is our justification for the present short-period analysis.

A recent study by Lotka and Spiegelman[11] provided us with data more suitable than any available earlier for such an analysis. These investigators computed maternity frequencies for the white population according to order of birth for the expanding Birth Registration Area of the United States by years from 1920 through 1938. "For this purpose, maternity frequencies have been computed separately for each order of birth within each quinquennial age group of women in the reproductive period. Then the frequencies were summed for all these quinquennial age groups so as to give a total for each order of birth, without regard to the age of the mother."[12] Thus, for the interwar period, "efficient" rates of births on the basis of those exposed to the risk of giving birth are available, whereas earlier studies had to depend on crude rates, the variations in which were partly due to extraneous variations in the base on which they were computed. Furthermore, the Lotka-Spiegelman series, being of the abstract type, are uninfluenced by variations in the proportions of women at different ages *within* the childbearing period. Finally, their organization of the data by order of birth makes possible more clear-cut analysis of the direct response than has been possible previously.

Lotka and Spiegelman did not smooth their series nor did they ana-

[10] *Ibid.*, pp. 99–100.

[11] Alfred J. Lotka and Mortimer Spiegelman, "The Trend of the Birth Rate by Age of Mother and Order of Birth," this JOURNAL, vol. 35, 1940, pp. 595–601.

[12] *Ibid.*, p. 596.

lyze the cyclical fluctuations. Their primary interest was in the general direction of change in the several orders of birth. They referred, however, to analyses which had been published in the *Statistical Bulletin of the Metropolitan Life Insurance Company* where first differences of birth rates and of marriage rates for New York State were correlated and showed that "almost without exception, an increase in the marriage rate from one year to the next was followed—with a lag of one year—by an increase in the birth rate of first children. When we recall the high correlation that has been found to exist between the marriage rate and economic conditions, it is to be expected that the time series of first births will also be closely correlated with economic conditions." They pointed out further that "the year 1933 marks an important turning point" for first births and also, but to a lesser degree, for second births, both of which exhibited a marked upswing as a response to improving economic conditions. They detected no such upswing in births of higher order. They suggested, therefore, that "since . . . the birth rate is being influenced in increasing degree by the maternity frequency of first births, it appears that economic conditions are playing a greater role than in the past in determining the course of our birth rate."[13]

There are, clearly, alternative hypotheses that should be tested in order to determine the role that economic conditions play. On the one hand, it might be inferred that the economic influence is entirely an indirect matter: economic conditions influence the marriage rate, the marriage rate influences the rate of first births, and variations in first births are sufficiently strong to influence the course of variations in total births. It is our contention that an hypothesis of this sort can be tested only if the divergent trends are first eliminated from each order of birth series and deviations from these trends are then correlated with deviations from the trend of an index of economic conditions. If the oscillations of birth orders higher than the first are uncorrelated with business cycles, this hypothesis will (within the limits set by the data and the techniques of analysis) be confirmed. If, however, these oscillations are significantly correlated with business cycles, the evidence may be considered to favor an alternative hypothesis, namely, that there is a direct response of births to business cycles, independent of the correlation of marriage rates with birth rates.

Through the courtesy of Dr. Lotka, the Metropolitan series were made available to us, and it is the purpose of this paper to explore these series with the same techniques used in analyzing the relationship of other series to business cycles. At the same time, in order to bring the

[13] *Ibid.*, p. 598.

Ogburn-Thomas analysis up to date, we have made a similar analysis of births by order and of marriages for a group of northeastern industrial states. We refer to these series subsequently as "Metropolitan series" and "our series."

A brief description of our series follows. *Marriage rates:* From compilations made by Stouffer and Spencer,[14] from scattered published reports of State Bureaus of Vital Statistics, and by correspondence, we were able to obtain the yearly number of marriages from 1919 to 1939 for New York, Connecticut, and Massachusetts; 1919–1938 for New Jersey; 1921–1937 for Pennsylvania; 1919–1937, and 1939 for Rhode Island. These were combined into a single series and rates per 1,000 estimated annual population computed, annual population being estimated by straight line interpolation between the censuses of 1920 and 1930, and between those of 1930 and 1940. *Crude birth rates:* The total number of births for each of the same states was obtained annually from 1919 to 1938 from *Vital Statistics of the United States*, and *Birth, Stillbirth, and Mortality Statistics*, and rates per 1,000 population similarly computed. *Birth rates by order of birth:* First, second, third, fourth, and fifth and subsequent births were obtained from the same source for all these states except Massachusetts, which does not collect data on order of birth. Rates were again computed per 1,000 population.

Our marriage series thus included one more state than our birth order series. Due, however, to the introduction of residence and medical requirements for marriage in several of these states, it was felt that the inclusion of Massachusetts data gave a fairer picture of variations in the marriage rate than if this series had included exactly the same states as used in the birth order series.

As noted, the Metropolitan series consists of maternity frequencies by order of birth. We used their data for first, second, third, and fourth births and combined their data for all orders higher than fourth into a series comparable with our "fifth and subsequent births." Our birth order series have one advantage over the comparable Metropolitan series: they are limited to a group of highly industrialized states and the same states are included throughout the period, whereas the Metropolitan series cover the expanding birth registration area. Since the states admitted to the area during the period under consideration were among the most rural in the nation, the Metropolitan series cannot be

[14] Samuel A. Stouffer and Lyle M. Spencer, "Marriage and Divorce in Recent Years," *Annals of the American Academy of Political and Social Science*, vol. 188, 1936, pp. 56–59.

Samuel A. Stouffer and Lyle M. Spencer, "Recent Increases in Marriage and Divorce," *American Journal of Sociology*, vol. 44, 1939, pp. 551–554.

considered as economically homogeneous. But the Metropolitan series are far superior to ours in another respect, namely, control of the factors of age and sex in the basic population.

As an index of business cycles, we used the United States Bureau of Labor Statistics adjusted index of factory employment, with the base period at 1923–1925.[15]

Our techniques of analysis may be briefly described as follows: Trends were computed for all series (ours, Metropolitan, and business cycles). Linear trends were found, upon trial, to be unsatisfactory. Second degree parabolas were, therefore, fitted by the least squares method, and were found to cut the cycles satisfactorily. Per cent deviations of data from trends were then correlated as follows: business cycles with marriages; business cycles with each order-of-birth series and with total birth series; marriages with our first-birth series. In all cases, birth series lagged by one year. Both linear and curvilinear (second degree parabolic) regressions were computed, as well as the net regression of first births (our series) on business cycles holding the marriage rate constant. Similarly, both the Pearsonian correlation coefficient (r) and the correlation index (ρ) were computed for each series. The relationships were, in all cases, approximately linear (see Chart I). There was slight second degree curvilinearity in the relationship between business cycles and some of the higher orders of birth. The curvilinearity was, however, so slight and of such doubtful statistical significance, that we accepted the relationship as being linear. Analysis of our series and the Metropolitan order-of-birth series yielded highly consistent results.

The correlation between business cycles and total births for this interwar period was very much higher than any that has come to our attention for any country at any earlier period.[16] For the Metropolitan series, $r = +.80$; for our series, $+.79$. These coefficients are unquestionably indicative of a very close relationship between economic conditions and birth rates. How far can this be attributed merely to the underlying relationship between marriage rates and the business cycle, on the one hand, and marriage rates and birth rates on the other? To answer this question, we first tested the relationship by partial correlation analysis of our series for marriages and first births. Simple correlation analysis of

[15] *Monthly Labor Review* and *Federal Reserve Bulletin*, various issues since 1935.

[16] Correlation coefficients between the total birth series of each of the six states and the business cycle index were also computed. They are: Massachusetts, +.46; Rhode Island, +.58; Connecticut, +.79; New York, +.68; New Jersey, +.77; Pennsylvania, +.70. The damping down of the coefficients for individual states due to random fluctuations, is apparent, since the coefficient for the six states combined was +.79.

CHART I

RELATIONSHIP BETWEEN BUSINESS CYCLES 1919–1937 AND BIRTHS (ONE YEAR'S LAG) AS SHOWN BY SCATTER DIAGRAMS AND REGRESSIONS, INCLUDING NET REGRESSION OF FIRST BIRTHS, HOLDING MARRIAGE RATES CONSTANT

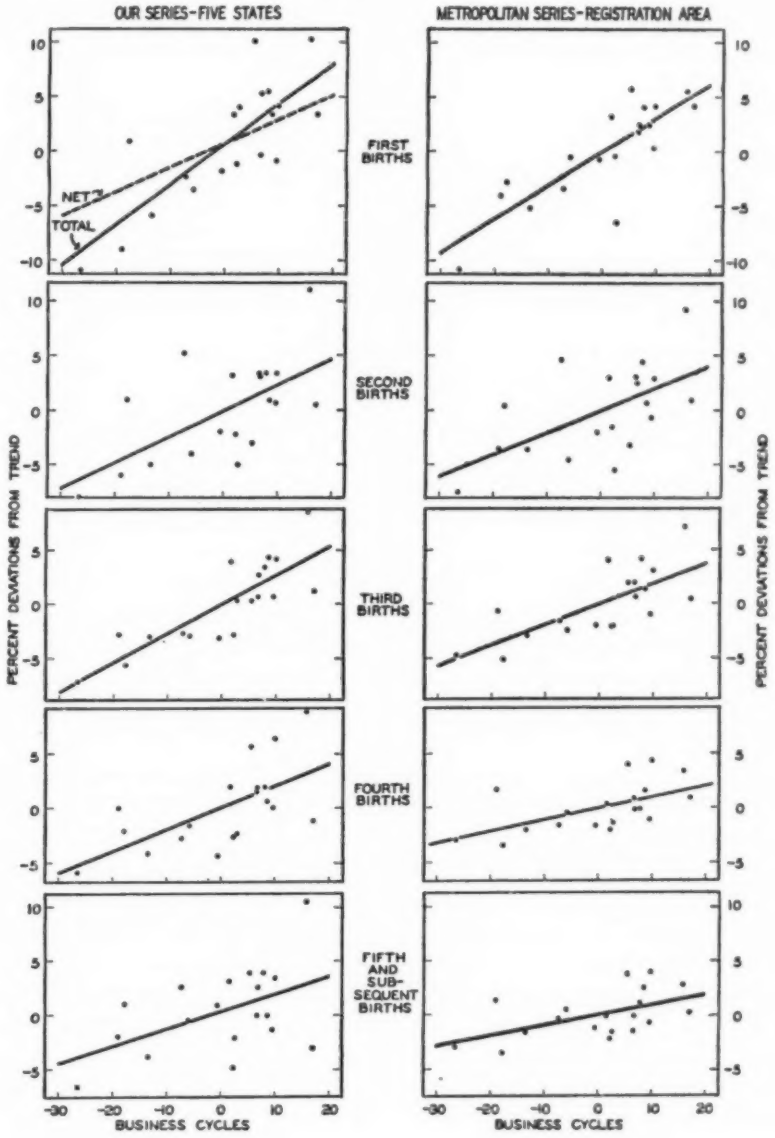


TABLE I

MEASURES OF RELATIONSHIP BETWEEN DEVIATIONS FROM TRENDS OF BUSINESS CYCLES, MARRIAGE RATES, AND BIRTH RATES, 1919–1937*

Independent variable (x)	Dependent variable (y)	r	ρ**	Linear regression equation	Standard error of estimate
	(Our series)				
Business cycles	Marriages	.848	.868	$y = +.363 + .460x$	3.38
	First births	.793	.795	$y = +.685 + .370x$	3.37
	Second births	.623	.626	$y = -.018 + .238x$	3.54
	Third births	.831	.837	$y = -.052 + .269x$	2.13
	Fourth births	.632	.646	$y = +.061 + .199x$	2.88
	Fifth and subsequent births	.493	.499	$y = +.381 + .158x$	3.30
Marriages	First births	.778	.781	$y = +.458 + .671x$	3.47
	(Metropolitan series)				
Business cycles	First births	.830	.832	$y = -.089 + .308x$	2.44
	Second births	.589	.593	$y = -.056 + .202x$	3.28
	Third births	.723	.732	$y = -.048 + .189x$	2.14
	Fourth births	.571	.595	$y = -.022 + .105x$	1.78
	Fifth and subsequent births	.531	.535	$y = -.007 + .092x$	1.75

Total linear regression, business cycles (x_1) and marriages (x_2) independent, and births (y) dependent.

$$y = .5675 + .222x_1 + .3245x_2$$

Net regression (business cycles and first births, holding marriages constant).

$$y = .7229 + .222x_1$$

* Birth rates lagging one year.

** Second-degree parabolic relationship.

business cycles with first births and marriages yielded coefficients of +.793 and +.848 respectively; while first births were themselves correlated with marriages by +.778. Multiple correlation analysis showed that when both business cycles and fluctuations in marriage rates were considered as independent variables and first births as the dependent variable, the coefficient was +.818. Taken in conjunction with the simple coefficients of +.793 and +.778 this result leads to the inference that the increase in "explainable" variance attributable to business cycles may be somewhat greater than the comparable increase attributable to marriage fluctuations. That is, when marriage fluctuations are held constant, the partial correlation between business cycles and first births is +.40; whereas where business cycles are held constant, the partial correlation between marriage rates and first births is +.33. Since the differences are of doubtful statistical significance, partial correlation analysis gave rather ambiguous results, and was further unsatisfactory because of the small number of observations and the inherent unreliability of our marriage series.

The question can be approached far more directly by analyzing the

correlation with business cycles of orders of birth higher than the first. For it is highly improbable that births of higher order than first are dependent to any extent upon marriages of the year preceding these births. Any observed relationship between business cycles and second, third, fourth, fifth, and subsequent births of the following year may, therefore, be assumed to indicate direct control and release of births in response to economic conditions.

From second through fifth and subsequent births, both our series and the Metropolitan show what might be called a very high to moderately high correlation with business cycles; for second births, +.62 and +.59, for third, +.83 and +.72; for fourth, +.63 and +.57; and even for fifth and subsequent, +.49 and +.53. The regression coefficients are, in all cases, unquestionably significant (see Chart I). The slope is steepest for first births and lessens rather systematically through subsequent orders. On the average, a change of 10 in the per cent deviations of business cycles from their trend is accompanied by the following changes in the same direction in our series: 3.7 per cent in first births; 2.4 per cent for second; 2.7 per cent for third; 2.0 per cent for fourth; and 1.6 per cent for fifth and subsequent. The changes in the Metropolitan series are consistent in pattern, but slightly less in amount, namely: 3.1 per cent, 2.0 per cent, 1.9 per cent, 1.1 per cent, and 0.9 per cent. Thus, although the correlation is marked, there is evidently a diminishing effect from first through subsequent orders. When, however, the net regression (for our series), of first births on business cycles, holding marriage rates constant, is computed, the slope is greatly diminished (from 3.7 per cent per 10 units to 2.2 per cent), becoming, in fact, less than those of second or third births and only slightly greater than that of fourth births. This suggests that the direct restriction of first births during depressions is somewhat less than that of immediately higher orders, a finding that is compatible with the spread of the small family system. The conclusion seems inescapable that, during the past two decades, the preventive check has been extensively in operation. Marriages are "controlled" during depressions, and, within marriages, births of all orders are likewise controlled. Since birth rates of higher orders are overweighted with births to the lower income, occupational, and educational classes, it is clear that the birth control movement has penetrated deep into the social structure during the past two decades.

Examination of Chart II throws further light on the nature of the relationships. The depression of the early twenties is clearly reflected in all orders except second and fifth and subsequent; while the depression and subsequent revival of the late thirties leave their marks on all or-

CHART II

RELATIONSHIP BETWEEN BUSINESS CYCLES AND MARRIAGES AND BIRTHS (ONE YEAR'S LAG) AS SHOWN BY TIME SERIES OF PER CENT DEVIATIONS FROM TRENDS (IN UNITS OF STANDARD DEVIATIONS)

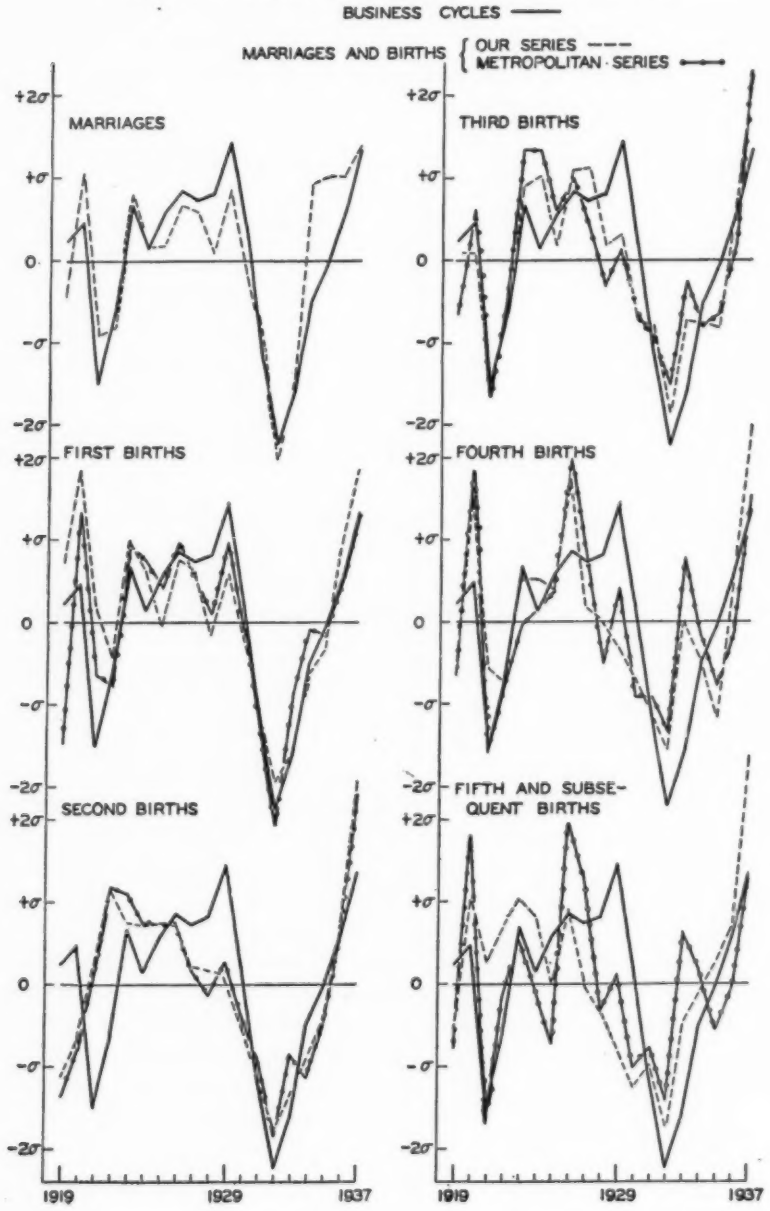


TABLE II

BUSINESS CYCLES INDEX, BIRTH RATES BY ORDER OF BIRTH, FIVE STATES; AND MARRIAGE RATES, SIX STATES

Year	Crude rates per 1,000 population							Per cent deviations from secular trends*						
	Business cycle index	Marriages	First births	Second births	Third births	Fourth births	Fifth and subsequent births	Business cycle index	Marriages	First births	Second births	Third births	Fourth births	Fifth and subsequent births
1919	106.7	9.8	5.6	5.2	3.7	2.6	6.1	2.7	− 2.9	4.0	−5.0	.3	−2.3	−2.1
1920	107.1	10.4	6.8	4.9	3.6	2.5	6.0	5.5	6.6	10.2	−3.0	.3	5.7	3.9
1921	82.0	8.9	7.1	4.9	3.5	2.6	6.1	−17.7	− 5.9	.9	1.0	−5.6	−2.1	1.0
1922	90.7	8.7	6.4	5.0	3.2	2.3	5.6	− 7.3	− 5.3	− 2.4	5.2	−2.7	−2.7	2.6
1923	103.8	9.4	6.1	5.1	3.2	2.2	5.5	7.9	5.1	5.5	3.4	3.4	1.9	3.9
1924	96.4	8.8	6.5	4.9	3.3	2.2	5.3	1.7	1.0	3.4	3.2	3.9	1.9	3.1
1925	99.8	8.6	6.3	4.8	3.2	2.1	5.0	6.7	1.1	− .3	3.3	.7	1.5	.0
1926	101.7	8.7	6.0	4.7	3.0	2.0	4.6	10.0	4.3	4.2	3.4	4.2	6.4	3.4
1927	99.5	8.5	6.2	4.6	3.0	2.0	4.5	8.6	3.8	3.4	.9	4.3	.6	.0
1928	99.7	8.1	6.1	4.4	2.9	1.8	4.1	9.6	.5	− .9	.7	.7	.0	−1.3
1929	106.0	8.4	5.8	4.3	2.7	1.7	3.8	17.1	5.5	3.3	.5	1.2	−1.2	−3.0
1930	92.4	7.8	6.0	4.2	2.6	1.6	3.5	2.4	− 1.1	− 1.2	−2.2	−2.8	−2.6	−4.8
1931	78.1	7.3	5.7	4.0	2.4	1.5	3.2	−13.4	− 6.9	− 5.9	−5.0	−3.0	−4.1	−3.8
1932	66.3	6.6	5.4	3.8	2.3	1.4	3.0	−26.5	−15.5	−10.8	−7.9	−7.1	−5.8	−6.6
1933	73.4	7.1	5.1	3.6	2.1	1.3	2.7	−18.9	− 9.2	− 8.9	−6.0	−2.8	.0	−1.9
1934	85.7	8.3	5.2	3.6	2.1	1.3	2.6	− 5.9	5.9	− 3.5	−4.0	−2.9	−1.6	− .4
1935	91.3	8.4	5.5	3.6	2.0	1.2	2.4	− .5	6.5	− 1.8	−1.9	−3.1	−4.3	.9
1936	99.0	8.5	5.6	3.6	1.9	1.1	2.2	6.9	6.6	5.3	3.1	2.7	1.9	2.6
1937	108.6	8.8	6.0	3.7	1.9	1.1	2.0	15.9	9.0	10.3	11.1	8.6	8.9	10.5
1938	89.7	7.7	6.3	3.9	1.9	1.1	1.9							
1939	96.9	8.1												

Constants of Secular Trends

	Period	Origin	*a*	*b*	*c*	Standard deviations of per cent deviations from secular trends
Business cycles	1919–1939	1929	90.50	−.37	+.0965	11.81
Marriages	1919–1939	1928–29	7.96	−.09	+.0126	6.40
First births	1919–1938	1928–29	5.87	−.05	+.0035	5.52
Second births	1919–1938	1928–29	4.31	−.09	+.0009	4.52
Third births	1919–1938	1928–29	2.73	−.10	−.0001	3.83
Fourth births	1919–1938	1928–29	1.75	−.09	+.0010	3.72
Fifth and subsequent births	1919–1938	1928–29	3.98	−.25	+.0009	3.80

* Birth rates lagging by one year.

ders. Third and later orders, however, show a sharper revival in 1934 and more of a recession in 1935–36 than the index of employment for 1933 and 1934–35 would lead one to expect. There is, of course, always a possibility that bad trend fitting in one or the other series distorts the relationship. The failure of second births to follow the business cycle index during most of the twenties cannot be readily explained, especially since third and even fourth births responded so well. The lack of correspondence may, of course, be due in part to the aftereffects of the exceptionally high marriage cohorts of 1918. The post-war inflation of marriages would probably have had its primary effect on first births in 1919, which is not included in our series. By 1921–22, these cohorts might well have had a sufficient number of second children to affect this series. That the rhythm extended beyond the second child is, however, highly doubtful.

Table II gives the basic data and the statistical constants for our own series.

A SEQUENCE OF HISTORICAL RANDOM EVENTS: DO JESUITS DIE IN THREE'S?

By J. Solterer
Georgetown University

Amongst the men of the Society of Jesus there exists a persuasion that they die commonly in groups of three. Such tradition or as some would call it, superstition, is not limited to this Society nor to grouping of deaths alone; in fact, there are even proverbs in different languages which point to relatively high frequencies of groups of three within series of historical random events, or at least to a prevalent tendency toward bunching in such sequences.

The Rev. E. C. Phillips, S. J., of Georgetown University examined in this respect[1] the 597 deaths which had occurred in this century (1900–1939 inclusive) in the Maryland-New York Province of the Society of Jesus. These forty years, including nine leap years, contain 14,609 days, spanned by 596 intervals between successive deaths; hence the mean interval between two successive deaths is 24.5 days and 49.0 days between the first and third of a group of three deaths.

The original data upon which this discussion is based, were taken from the official annual necrologies, issued for private circulation among the members of the Jesuit order. These lists give only the calendar date of each death, and hence the intervals could not be determined in units of less than one whole day.

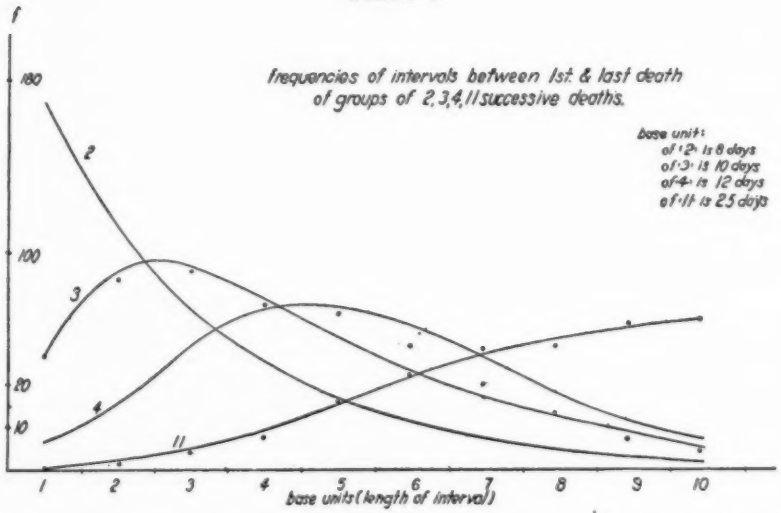
Father Phillips found that there were 68 groups of three deaths occurring in intervals of 17 days (i.e. in one third of the mean interval) or less, and 28 further groups of three within 18 to 25 days; a total of 96 groups of three deaths which occurred in intervals equal to or shorter than one half the mean interval of 49 days. In some of these short intervals there were more than three deaths following each other. This was particularly noticeable during an epidemic of flu when there were ten deaths in 31 days, from September 26 to October 26, 1918.

Father Phillips concluded that the persuasion in question has a solid foundation in fact. Out of the total of 597 deaths, containing 199 triplets, there are in fact 96 such close groups whence it follows that more than 48 per cent of all deaths occurred in groups spanning at most 25 days. The fact of grouping was taken to be established but the nature of the fact was left conjectural with the added suggestion that the phenomenon was one of chance.

[1] Rev. E. C. Phillips, S. J., "Do Deaths Come in Groups of Three," *Bulletin of the American Association of Jesuit Scientists* (Eastern Section), Loyola College, Baltimore, Md., vol. XVII, no. 4, pp. 196–199 (May 1940).

Rev. J. T. O'Callahan, S. J., of Holy Cross College, presented a further investigation to the National Science Convention of the American Association of Jesuit Scientists.[2] He arranged the time intervals elapsing between the first and last death of groups of two, three, four and eleven deaths in frequency distributions and showed that they constitute a related series of the general statistical curve, Type III of Pearson. In the case of doublets the curve is L-shaped, the shortest doublets having the highest frequency. The larger the group, the less does the frequency distribution depart from the approximately normal

CHART I



curve, observed in the case of groups of eleven deaths. These curves, adapted from an unpublished diagram prepared by Father Phillips, are shown in Chart I. This seems to be the first published concrete case in which the frequency distribution of a single series of observations changes from an extreme non-normal form to a normal form as the unit of grouping is varied.

Father O'Callahan concluded from the goodness of fit that the skewness of the distribution, i.e. the prevalence of short intervals or the grouping effect, is due to chance and is characteristic of all random series of historical events. The frequency of observed short interval quadruplets is relatively so small as not to attract attention. The short interval doublets are very numerous but they very frequently combine

[2] Rev. J. T. O'Callahan, S. J., "Deaths of Jesuits in Groups of Three; The General Grouping of Random Historical Events," *Bulletin*, vol. XVIII, no. 1, page 52, (October 1940).

in observation with the next death and so form a short interval triplet. Thus, the random series gives rise to the persuasion that events occur in groups of three only.

This analysis rests essentially on the notion of goodness of fit of empirical curves to observed data and, furthermore, on the fact that large groups falling within short intervals are scarce. The test of goodness of fit, however, is not sufficient to establish the nature of the function required. At best, such tests can establish the randomness of the deviations from the curve. What is needed first is an hypothesis on general theoretical grounds; this may later be tested for goodness of fit.

The higher frequency of closely spaced deaths when compared with more widely spaced ones is to be expected in case each death is a random event. If we chose a sufficiently short observation interval, the probability of a certain death falling within it is less than that of falling without. Consequently, the distribution of deaths per short observation interval must be skewed. For an extremely short interval this distribution will be L-shaped, the distribution of rare events. If the observation interval is indefinitely long the normal distribution will be approached. The impression we get from the observation of a historical random series obviously depends on the length of the observation interval. A long interval will leave us with the conviction that deaths are more or less evenly spaced, a short one that they occur in spurts or particular, characteristic groups.

For each observation interval there will be a particular modal value for deaths per interval. If among all possible intervals for any series of observations there would be one, enforced by Nature on all observers alike, then the respective modal value for deaths per interval would determine what grouping effect we observe.

It is quite clear that for many fortuitous time series this observation interval is habitually or arbitrarily short. The result is the impression of bunching. Series of automobile accidents per week or month in our cities, fire hazards and the like are good examples. The influence on bunching of profits of the variation of the accounting or fiscal period may also be mentioned in this connection. Actually, many governments have discovered empirically that it is possible to reduce the irregularity of some of their revenues by increasing the corresponding fiscal period. Of course, it is not asserted that such bunching might not also be the consequence of systematic causes; nevertheless, an old rule of wisdom advises us to look at events *sub specie aeternitatis*.

The distribution function of deaths per observation interval is the same as that of the Brownian motion without the so-called probability

after effect.[3] The usual way of observing the Brownian motion consists of fixing the attention on a square formed by a net of mutually perpendicular, equidistant lines on the bottom of a shallow vessel, containing a colloidal solution, and counting the particles contained within this square from second to second or in any other chosen interval. Our vessel is the Maryland-New York Province, our solution consists of all men in this province, and our particles within the square are the deaths occurring in this province. We neglect the following five factors: (1) the flu epidemic 1918–1919; (2) the variation in the population in the province from year to year during the forty years considered; (3) the change (increase) in life expectancy during this time; (4) the difference in length of months; (5) the seasonality of mortality. The Chi-square test employed later justifies this neglect.

The Brownian probability after-effect results from the relation between the average velocity of particles and the size of the square as well as the observation interval. Only small changes in the number of particles in the square can be expected to occur in intervals not large relative to the velocity. In our case, however, the number of deaths in any one interval is independent of the number in any other, with the exceptions just noted above.

If all the observed deaths, numbering N, were equally spaced throughout all time T, a certain number ν would be found in each observational interval t. This is the mean number of deaths per observation interval.

$$\frac{\nu}{N} = \frac{t}{T} \quad \text{or} \quad \nu = N\frac{t}{T} \tag{1}$$

The probability of a certain death falling within t is: $P_t = t/T$.
The probability of a certain death not falling within t is:

$$P_{T-t} = \frac{T-t}{T}.$$

The probability of n deaths falling within t is:

$$P_{n(t)} = \frac{N!}{n!(N-n)!}\left(\frac{t}{T}\right)^n\left(\frac{T-t}{T}\right)^{N-n}.$$

Substituting ν from (1) we get:

$$P_{n(t)} = \frac{N!}{n!(N-n)!}\left(\frac{\nu}{N}\right)^n\left(\frac{N-\nu}{N}\right)^{N-n}$$

[3] Cf. Von Mises, *Probability, Statistics and Truth*, pp. 261, 262.

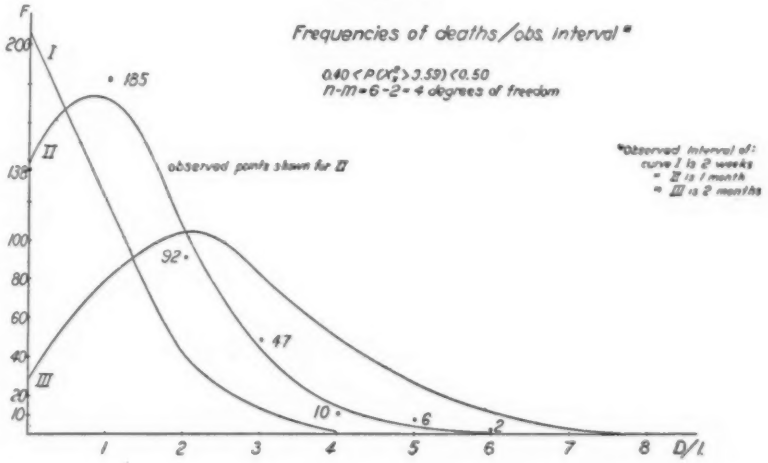
and for large N we get Poisson's distribution:

$$P_{n(l)} = \frac{\nu^n}{n!} e^{-\nu}.$$

Chart II gives the respective curves for the observation intervals of two weeks, one month and two months. The corresponding ν's are:

$$\nu_{\mathrm{I}} = 0.6 \text{ death}$$
$$\nu_{\mathrm{II}} = 1.2 \text{ deaths}$$
$$\nu_{\mathrm{III}} = 2.5 \text{ deaths.}$$

CHART II



These curves, similar to those used by Father O'Callahan tell us indirectly that successive deaths occur in differently spaced groups since the empty or sparsely populated intervals are of ever greater frequency the smaller the observation interval. However, in both procedures the grouping effect depends on an arbitrary choice of either the size of the observation interval or the size of the observed group.

Popular observations, supporting the belief of deaths in three's, are independent of such choices since they may take any number of deaths as a group and, also, need not use any constant observation interval. The procedure will be to begin observation with any death, observe intervals between successive deaths, and derive from their relative lengths the impression of a pattern of the sequence. Correspondingly, we need the distribution of intervals between successive deaths. It will then be easy to determine the probability of any combination of different successive intervals because each interval is an independent event.

The function required has been derived by physicists in connection with their study of emission of α particles (scintillation).[4]

The different intervals observed between two successive deaths may be thought of being made up of a number of equal elementary intervals following each other. These elementary intervals are so small that none of them contains more than one death. If there are M such intervals and μ is the probability that a particular interval contains one death, there will be $M\mu$ deaths in all, and $M/M\mu = \lambda$ will be the average interval between two successive deaths. μ and λ are reciprocal. The probability that, starting from a certain death, there will be a series of τ elementary intervals without any death, followed by a death within a very small interval $\partial\tau$ is

$$(1 - \mu)^{\tau}.\mu.\partial\tau = \left(1 - \frac{1}{\lambda}\right)^{\tau} \frac{1}{\lambda}\,\partial\tau.$$

If $\mu \to 0$, $\lambda \to \infty$. τ and $\partial\tau$ are numbers of elementary intervals and approach ∞ also. The ratios τ/λ, $\partial\tau/\lambda$, however, remain finite and are independent of the unit in which τ and λ are measured. Then the required probability will be:

$$\frac{1}{\lambda}\, e^{-\tau/\lambda}. \tag{2}$$

From this expression we see that the shorter the interval τ, the more frequent its occurrence. The preponderance of short intervals leads to the group impression. The latter is intensified if λ is large, i.e. if the mean number of deaths occurring in any fixed interval is small. We know this already from our previous discussion.

From (2) can be deduced an expression for the probability of an interval between two successive deaths falling within a particular finite range of values of τ. Integrating (2), the probability of an interval being between τ_1 and τ_2 is:

$$\int_{\tau_1}^{\tau_2} \frac{1}{\lambda}\, e^{-\tau/\lambda} d\tau.$$

$$\pi = (1 - e^{\tau/\lambda}) \tag{3}$$

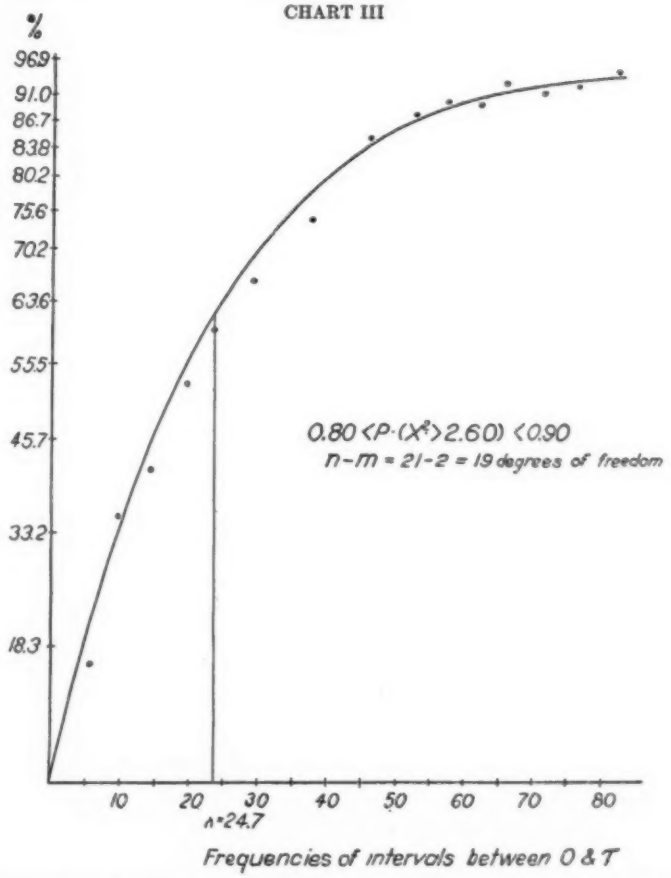
gives us the probability of an interval between zero and τ. Chart III shows this function and the corresponding observed values.

Any grouping of intervals presupposes classification into either short or long intervals, or intervals longer or shorter than the preceding one. If the latter classification be chosen, we note the following:

The total frequencies of intervals, shorter and longer than the preceding interval are equal because the two respective probabilities are

[4] Cf. W. N. Bond, *Probability and Random Errors*, p. 19.

CHART III



complementary for each and all intervals. Of course, there will be also longer series of shorter in succession, or longer in succession, or combinations thereof. These will also be equally numerous for the same reason as given in the case of shorter or longer taken singly. In the present series there were observed:

Shorter	301	Shorter-longer	202
Longer	287	Longer-shorter	200
Equal	7		
Shorter-longer-longer-shorter			44
Longer-shorter-shorter-longer			56
Shorter-3 successive longer-shorter			14
Longer-3 successive shorter-longer			11
Shorter-4 successive longer-shorter			2
Longer-4 successive shorter-longer			2

This classification, however, does not lead us to any outstanding group impression as the one we are attempting to explain. We turn to the second of the classifications mentioned before, viz. classification of intervals into short and long. This requires a definition of short and long. Our function (3) has only one parameter, λ. Choosing any other value as the dividing value between short and long will introduce arbitrary elements. If, for instance, we should take the median value (17 days) of the interval series for this purpose, the chances would be equal that any interval, whether long or short, would be followed by either short or long. It would be observed very soon that by adopting such classification, intervals would be called long which are only slightly longer than the median and yet occur very frequently. So the norm interval would be lengthened. If an interval longer than the mean should be taken as the norm, the opposite result would ensue. The only classification free from this difficulty is the one which calls any interval short which is less than the mean interval (25 days), and long any interval more than that.

From function (3) we notice that any such short interval is followed by another short one in about 64 cases out of a hundred, because successive intervals are independent events. The observer of the interval series will not only be impressed by the predominance of short intervals but also by the fact that any short interval tends, in the majority of cases, to be followed by another short interval. He will notice the pattern of the death triplet.

The probabilities of any other larger combination may be derived from the expansion $(p_s+p_l)^n$ where p_s, p_l are 0.64 and 0.36, and n the number of intervals in the desired group. The respective pure short combinations are always the most frequent. This fact extends the impression of bunching from triplets to even larger groups.

In the present case there were 377 shorts (375 required theoretically), and 258 groups of two successive shorts (240 required theoretically).

We may conclude that the persuasion in question, a supposed superstition, that Jesuits die in three's is confirmed by the series observed and, furthermore, has been shown to be the recognition of an important trait of all random historical sequences.

RELATION BETWEEN SEASONAL AMPLITUDES AND THE LEVEL OF PRODUCTION—AN APPLICATION TO THE PRODUCTION OF STEEL INGOTS

By Gustave S. Juliber[1]

There are three possible relationships between the business cycle and seasonal amplitudes of production. A particular relationship will depend on the circumstances surrounding an individual product.

In the case of steel, for example, the amplitudes of seasonal fluctuations tend to become larger as the general demand for that product declines. This tendency is entirely reasonable in view of modern business practices. During periods of recession business men often resort to hand-to-mouth buying; they do not build up stocks; they are doubtful of the future. Production, therefore, is largely seasonal in nature, and demand varies widely from month to month. During the upswing of the cycle seasonal fluctuations assume a different character. Demand increases over an extended period of time; business men become increasingly optimistic; they build stocks; the policy of hand-to-mouth buying is discarded. Because of the *continued* increase in demand, production is largely *cyclical* in nature, and month-to-month variations diminish as the peak of the cycle is approached. Moreover, seasonal amplitudes of steel production must decrease because of the fixed physical capacity which is present at a given point of time.

A second kind of relationship may exist. In the case of certain luxury goods, for example, seasonal amplitudes of production increase as the peak of the business cycle is approached. Here the circumstances are different, for the sale of these goods depends, in a large measure, on the purchasing power of consumers. When purchasing power is low or declining the consumer does not easily part with his money, either in or out of season. On the other hand, when the consumer's pockets are well filled his purchases in season are large.

Finally, there are some types of products whose seasonal amplitudes of production are but little affected, if at all, by general business conditions. A case in point is to be found in the motion picture industry, where the number of domestic feature-length films produced varies but little between the years of prosperity and the years of depression.

General recognition of the fact that seasonal patterns change in response to one or more causal factors has resulted in newer and more

[1] The author is indebted to Professor Olin W. Blackett and to Mr. Everett H. Johnson, both of the School of Business Administration, University of Michigan, for their helpful suggestions. None of them, however, is to be held responsible for the statements made in the text.

improved methods of determining seasonal variations.[2] In practice, however, seasonal changes are often assumed to take place only in relation to the time element, without regard to whether the business cycle is reaching peak levels or approaching the depths of depression. The latter factor is important and it warrants greater attention than it has thus far received.[3]

Kuznets is one of several persons who have considered this problem.[4] His Amplitude Ratio,[5] however, does not measure the relationship between seasonal amplitude and the level of business activity directly (i.e., month by month). Moreover, he assumes an average seasonal amplitude for a continuous period of time. This is not consistent with the tendencies previously described, for the period selected may include any or all phases of the business cycle.

The purpose of this paper is to suggest a more direct method of attacking this problem. The case at hand concerns the production of steel ingots. The method does not assume an average seasonal amplitude in relation to time. In fact the time element, as such, is unimportant.

How can the general level of steel ingot production be measured? A simple and direct method is to express production in terms of per cent of capacity. Thus, when production takes place at, say, 95 per cent of capacity, the industry is operating at or near the peak of its cycle. Conversely, if the productive rate is, say, 40 per cent, then the industry is operating at or near the low point of its cycle.

Although per cent of capacity is a somewhat imperfect indicator of absolute output (because increases in capacity are not taken into account), the use of this measure has several characteristics to recommend it. In the first place, seasonal amplitudes must diminish as absolute production reaches peak physical capacity. Secondly, the rate of production does reflect the movements of the business cycle regardless of capacity increases. In the third place, per cent of capacity is readily understood by the business man. Finally, the use of this measure of steel activity permits the interpolation of average seasonal amplitudes at rates of production for which data are not available.

[2] For a summary of both the older and the newer methods, and an evaluation of them, see "Methods of Computing and Eliminating Changing Seasonal Fluctuations," by Horst Mendershausen, *Econometrica*, Vol. 5, No. 3 (July 1937) pp. 234–262.

[3] The writer does not mean to imply that the matter has escaped notice. Kuznets, Wald, and others, have attacked the problem. However, many published indices of production do not take this factor into account. Even the majority of text books on general statistics do not treat this matter directly, but discuss at length the methods of determining changing seasonals in relation to the temporal factor.

[4] *Cf. Seasonal Variations in Industry and Trade*, by Simon Kuznets, National Bureau of Economic Research, New York, 1933, Chapters X and XI.

[5] *Ibid.*, p. 324.

The actual measurement of the relationship between average seasonal amplitudes and percentages of capacity is similar in method to that of determining moving seasonals with respect to the time element. The procedure is as follows:

1. Ratios to an adjusted moving average of absolute production data were obtained. In this instance a two-month moving average of a twelve-month moving average was used. The calculated average was plotted against the production data, then adjusted by hand in order to follow more closely the path of production, particularly at the turning points of the production cycles.

TABLE I
SUMMARY OF GROUPING PROCESS

(1) Group code	(2) Annual per cent of capacity	(3) Number of years	(4) Average capacity of years in column (3) (per cent)
—	Under 20	1	19.7*
—	20 Under 30	0	—
A	30 Under 40	5	37.3
B	40 Under 50	2	48.5
—	50 Under 60	0	—
C	60 Under 70	8	66.4
D	70 Under 80	6	74.4
E	80 Under 85	5	82.5
F	85 Under 90	5	86.7
G	90 Under 100	5	95.4
—	100 Over	1	101.6*

* Omitted. See text.

2. The adjusted ratios for each calendar year were then grouped according to the annual per cent of productive capacity in the corresponding year. Columns 1, 2, and 3 of Table I summarize the process. Any successive twelve-month periods for which capacities are known can be used instead of calendar years.

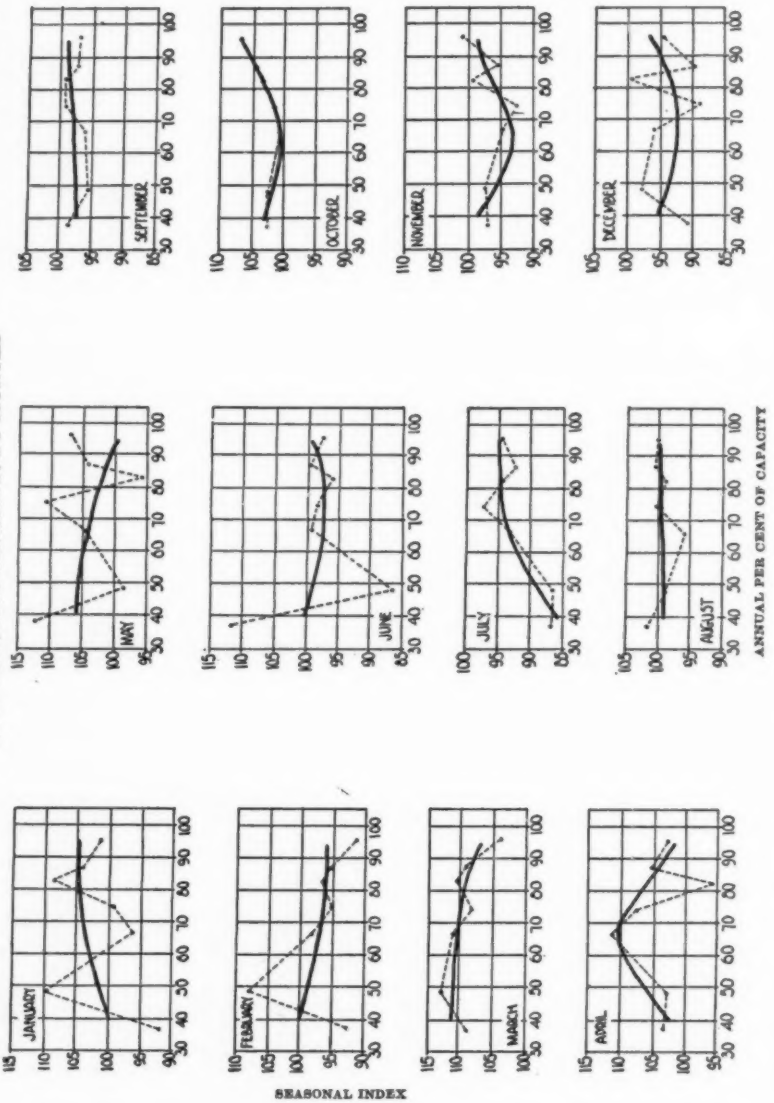
3. Seasonal variations were then calculated separately for each group in the usual manner. In this instance modified means were employed for all groups except Group B; in the latter case a simple mean was used.

4. Twelve scatter diagrams, one for each month, were then constructed. These are shown in Chart I. The vertical scale represents values of the seasonal index already obtained, while the per cent of capacity values are found along the horizontal axis. Each monthly index figure was plotted against the average per cent of capacity of its own group (given in Column 4, Table I). For example, the January index value for Group A was plotted at 37.3 per cent of capacity; the

January index value for Group B was plotted at 48.5 per cent of capacity. The other months were treated in similar fashion.[6]

CHART I

AVERAGE RELATION BETWEEN SEASONAL INDEX OF STEEL INGOT PRODUCTION AND ANNUAL PER CENT OF PRODUCTIVE CAPACITY: BY MONTHS



5. A line of average relationship was then drawn through the points by hand (see Chart I). The sum of the twelve curves at a given point

[6] It should be noted that the seasonal index which is finally obtained is related to percentages of capacity, and *not* to absolute units of production.

of capacity approximates 1200. Readings from these curves are given in Table II.

TABLE II

ESTIMATED AVERAGE SEASONAL INDEX OF STEEL INGOT PRODUCTION AT GIVEN ANNUAL PERCENTAGES OF PRODUCTIVE CAPACITY*

Month	Annual per cent of capacity					
	40	45	50	55	60	65
January	100.00	100.75	101.25	102.25	102.75	103.25
February	100.00	99.50	99.25	98.75	98.25	97.50
March	111.50	111.50	111.50	111.25	110.75	110.50
April	101.75	103.75	105.50	107.50	109.00	110.25
May	106.00	106.00	105.75	105.50	105.00	104.50
June	100.75	100.00	99.25	98.00	98.00	97.75
July	85.75	87.50	89.25	90.75	92.00	93.25
August	99.00	99.25	99.25	99.50	99.50	99.50
September	97.25	97.50	97.50	97.75	97.75	98.00
October	103.00	102.25	101.50	101.00	100.50	100.50
November	98.50	97.00	96.75	94.50	93.75	93.50
December	95.50	94.50	93.75	93.25	92.75	92.50

Month	Annual per cent of capacity					
	70	75	80	85	90	95
January	103.75	104.00	104.50	104.75	104.75	104.75
February	96.75	96.25	96.00	96.00	96.00	96.00
March	110.00	109.75	109.00	108.00	107.00	105.50
April	110.00	108.50	107.00	105.00	103.50	101.50
May	104.00	103.50	102.75	101.50	100.50	99.50
June	97.50	97.25	97.50	97.75	98.50	99.50
July	94.00	94.50	94.75	94.75	94.75	94.75
August	99.75	99.75	99.75	99.75	100.00	100.00
September	98.00	98.25	98.50	98.75	98.75	99.00
October	100.75	101.50	102.75	103.75	105.00	106.25
November	93.75	94.50	95.50	96.75	97.75	98.50
December	92.50	92.75	93.25	94.00	95.00	96.50

* Source: Readings from Chart I. Data are not adjusted for number of working days per month.

6. At each point of capacity, the twelve monthly deviations from 100 were determined. Three series of deviations were obtained: the average deviation for each calendar year; the deviation of the highest month in that period; the deviation of the lowest month in that period. These data are given in Table III, and also presented in Chart II.

The accuracy of the data presented in the charts and tables must be interpreted in the light of the limited amount of original data available for this type of analysis. The original data covered the period 1901 to 1940. The year 1916 (operating at 101.6 per cent of capacity) and the year 1932 (operating at 19.7 per cent of capacity) were excluded because these years stood alone and could not be grouped with similar

CHART II

AMPLITUDES OF ESTIMATED SEASONAL INDEXES OF STEEL INGOT PRODUCTION AT GIVEN ANNUAL PERCENTAGES OF PRODUCTIVE CAPACITY

(Source of data: Table III)

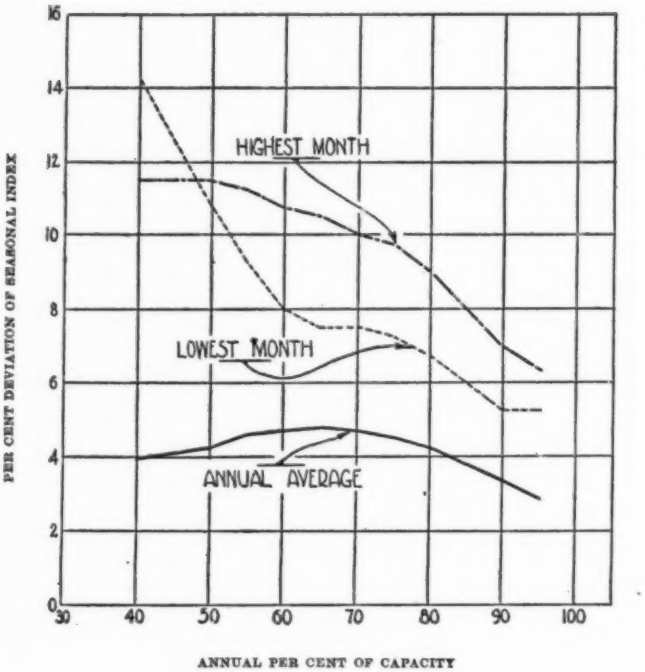


TABLE III

COMPARISON OF AMPLITUDES OF ESTIMATED SEASONAL VARIATIONS AND ANNUAL RATES OF STEEL INGOT PRODUCTION

Annual average capacity (per cent)	Deviation of highest month from 100 per cent	Deviation of lowest month from 100 per cent	Average annual deviation (per cent)
40	+11.50	−14.25	3.92
45	+11.50	−12.50	4.08
50	+11.50	−10.75	4.21
55	+11.25	− 9.25	4.58
60	+10.75	− 8.00	4.67
65	+10.50	− 7.50	4.75
70	+10.00	− 7.50	4.69
75	+ 9.75	− 7.25	4.50
80	+ 9.00	− 6.75	4.23
85	+ 8.00	− 6.00	3.77
90	+ 7.00	− 5.25	3.33
95	+ 6.25	− 5.25	2.85

years. In spite of the paucity of data, however, the results obtained are in agreement with the theory discussed previously.

Seasonal amplitudes have been here measured by two methods. The first method is that of obtaining seasonal deviations of the highest and lowest months from 100 for each calendar year with respect to the rate of production in that year. Amplitudes, when measured in this way, decrease as production approaches peak capacity, and increase as lower and lower levels of capacity are approached.

It is interesting to note that, at the lowest levels for which seasonals could reasonably be interpolated, the lowest months deviate more than the highest months until approximately 48 per cent of capacity is reached. The reason for this is probably to be found in the fact that the demand for steel is derived from the demand for steel products. When there is little demand for the products which contain steel, the latter will not be produced in any great quantity during at least several of the depression months.[7]

It is also interesting to note that the lines of maximum deviations in Chart II tend to meet at some point beyond 95 per cent of capacity. This is explained by the fact that the seasonal pattern of production would take the form of a straight line at a point slightly beyond 100 per cent of practical capacity if the industry were to operate at peak physical capacity in each month of the year.[8] Obviously, the deviations at this point would equal zero. Practically, however, a deviation of zero is not to be expected because some furnaces are always taken out of production for repairs, relinings, etc.

The second method of measuring seasonal amplitudes is that of obtaining the average seasonal deviation of the twelve months of each calendar year from 100 with respect to the rate of production in that year. As indicated previously, a deviation of zero is theoretically expected when operations take place at peak capacity during the year. Conversely, if operations were to cease altogether, then a "seasonal deviation" of zero would also be obtained, and the "seasonal pattern" in this theoretical case would also be a straight line. Between these two extremes, at some point of capacity, a maximum average amplitude will be found.

[7] This result may, however, be somewhat influenced by the moving average, which is particularly difficult to adjust by inspection at the severe troughs. If the moving average does not swing sufficiently into the troughs (i.e., the moving average is too high relative to the original data), then the very lowest months of actual steel output will yield an unduly small ratio to the average, so that the lowest months will deviate more than the highest months. The "goodness of fit" of the moving average used in this study was approved by competent persons.

[8] For relatively short periods the industry can operate at more than 100 per cent of practical capacity.

The course of the line of annual average deviations shown in Chart II is in harmony with these expectations. Between 40 and 65 per cent of capacity the average deviations increase; between 65 and 95 per cent of capacity the deviations decline. Just as a deviation of zero will never be actually reached at the very highest levels of production, a deviation of zero will never be actually reached at the very lowest levels, for the manufacture of steel is a continuous process.

The point of maximum average annual amplitude appears to occur at 65 per cent of productive capacity. Between 65 and 95 per cent of capacity the annual average deviations decline because of the steadily increasing pressure of physical capacity on steel production. Between 65 and 40 per cent of capacity, physical capacity exerts less and less influence on seasonal amplitudes.

The relationship between seasonal amplitudes of steel ingot production and the production cycle appears to be this: when production is approaching physical capacity seasonal amplitudes diminish because of the pressure of fixed capacity; when operations are approaching low levels seasonal amplitudes increase. Physical capacity apparently begins to exert its pressure on seasonal amplitudes of production at about 65 per cent of capacity. Prior to that point capacity apparently has little effect on seasonal amplitudes.

The seasonal variations of production for all products are not expected to be affected in the same manner by changes in the business cycle. It is reasonable to believe that two other relationships are possible, depending on the nature of the product in question. In some cases seasonal variations may be expected to increase in amplitude as the peak of the cycle is approached. The relationship in these instances would be inverse to that obtaining in the production of steel ingots. Finally, there are some products whose seasonal movements of production may be expected to be unaffected by changes in the business cycle.

It should be noted that cyclical movements are only one of a series of factors which may systematically alter seasonal amplitudes of production. The nature of the product in question will dictate further refinements.

PROBABILITIES IN LOGARITHMIC SKEWED DISTRIBUTIONS

By G. R. Davies and R. H. Smith
University of Iowa

The probabilities expressed in an ordinary table[1] of the normal curve, $z=(2\pi)^{-1/2}e^{-t^2/2}$, are commonly applied to the skewed distributions encountered in practice. It is obvious that this procedure may introduce inaccuracies, but little attention has been given to the degree of error. For example, to select the most elementary case, it is customary to say that if an item is drawn at random from a population whose mean and standard deviation are known, the probability is 0.95 that it will fall within a distance of 1.96σ from the mean; and this rule is applied even though considerable skewness characterizes the original distribution. Similarly, fiducial limits of a sample mean are commonly based on the assumed normal distribution of sample means; whereas, in fact, the skewness of the population is not eliminated in the means, but only reduced approximately by the factor $1/N^{1/2}$. It is the purpose of this paper to inquire into the degree of inaccuracy attributable to varying degree of skewness in stating such probabilities.

It will be assumed that the most practicable theoretical representation of the usual skewed curve is the so-called logarithmic normal distribution. Adequate descriptions of this curve from the standpoint of fitting it to data are available,[2] but for the purpose at hand a brief description of its relation to the normal curve will be in order. This description will deal with positive skewness only, since the conclusions sought may be equally well applied to either positive or negative skewness.

The logarithmic normal curve may most easily be described as an adaptation of the common binomial $(\frac{1}{2}+\frac{1}{2})^n$, the frequencies of which suggest a bell-shaped outline when plotted successively at the abscissae $X=0, 1, 2, \cdots n$. A corresponding logarithmic skewed binomial may be plotted by taking these abscissae as exponents of a suitable base; that is, by transforming them to $X=C^0, C^1, C^2, \cdots C^n$ where C is a constant greater than unity. When the binomial is so plotted, the frequency class intervals, relative to a central interval, will decrease serially

[1] The usual x/σ is here denoted, as in some published tables, by the more convenient symbol, t (cf. Rietz, H. L., *Handbook of Mathematical Statistics*, p. 209, Houghton, Mifflin and Company, New York, 1924).

[2] Croxton, F. E., and Cowden, D. J., *Applied General Statistics*, Prentice-Hall, New York, 1939, pp. 295–299. Davies, G. R., "The Logarithmic Curve of Distribution," this Journal, December, 1925, and "The Analysis of Frequency Distributions," *ibid.*, December, 1929.

toward the left and increase toward the right, and the height of the frequencies must be inversely adjusted so that the data may be accurately represented by graphic areas. That is, when the frequencies, plotted as rectangles, are transferred to the exponential scale, a relative contraction of the base must be offset by a proportionate expansion of the height, and vice versa.

If now a binomial so plotted in both its normal and skewed forms is changed from the discrete to the continuous X scale by making n approach infinity, then the normal curve and a related logarithmic skewed curve will be fairly well represented. If, further, the horizontal scale of the normal curve is "centered," with a mean of zero, and with the deviations written in σ units (x/σ or t), then the skewed curve will "center" at a geometric mean of 1, and its X scale will range between the limits 0 and ∞.

Before discussing the interrelationships of the two curves thus described, certain basic units of measurement should be considered. It is true that in practice the skewed curve is commonly measured by the arithmetic mean and the standard deviation, but for the purpose at hand the geometric mean and the so-called standard deviation ratio are preferable. The standard deviation ratio is not in common use but may be easily described by its relation to the geometric mean. If, for example, data conforming to the logarithmic normal curve were tabulated, and the mean and standard deviation were computed from the logarithms of X, obviously the antilogarithm of this mean would be the geometric mean, and similarly the antilogarithm of this standard deviation would be the standard deviation ratio. In practice the standard deviation ratio is not conveniently so computed, but its nature may be more readily understood from a consideration of such a method. Manifestly, both it and its logarithm, the logarithmic standard deviation to base 10 (log σ_r or $\sigma_{\log}$), constitute convenient indexes of the degree of skewness. By methods suggested later, the relation of σ_r to Bowley's quartile measure of skewness (Sk_Q) is found to be[3]

$$Sk_Q = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} = \frac{{\sigma_r}^{0.6745} - 1}{{\sigma_r}^{0.6745} + 1} \tag{1}$$

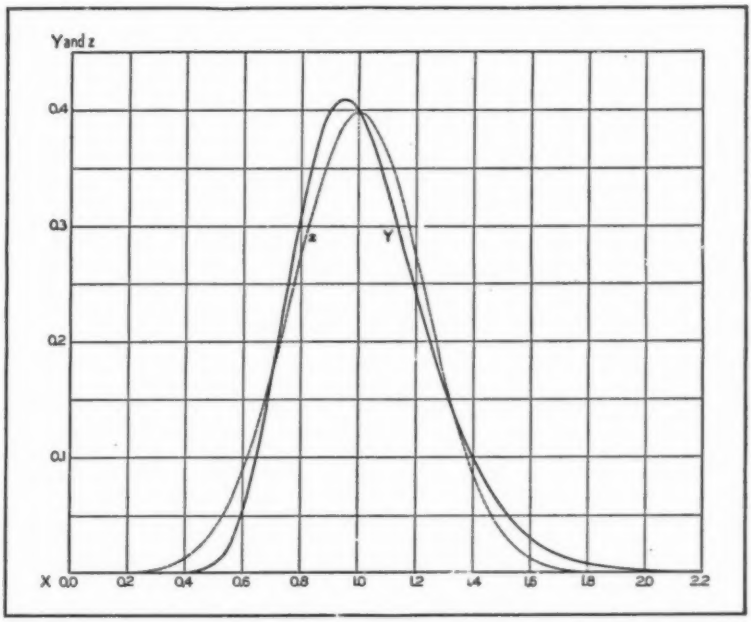
and

$$\sigma_r = \left(\frac{1 + Sk_Q}{1 - Sk_Q}\right)^{1.4826}. \tag{2}$$

[3] *Ibid.*, p. 253.

For convenience of exposition, a normal distribution and a related logarithmic normal distribution of the degree of skewness indicated by $\log \sigma_r = 0.1$ or $Sk_Q = 0.0775$ are plotted together on an arithmetic scale in the accompanying chart. A description of the method by which this is done will be in order.

LOGARITHMIC NORMAL CURVE, Y (SKEWNESS: $\log \sigma_r = 0.1$, $Sk_Q = 0.0775$), AND NORMAL CURVE, z, CENTERED AT $X = 1$; $\sigma = \log \sigma_r / 0.4343$



In constructing the chart, the normal curve given in the usual tables (z on x/σ or t) is taken as the starting point, and appropriate intervals for plotting—half x/σ or t units—are selected (see Table I). In order to avoid the use of two horizontal scales, however, and to make areas graphically comparable, the selected t points are converted to X_N positions by writing them as

$$X_N = 1 + \frac{t \log \sigma_r}{0.4343} \tag{3}$$

so that a standard deviation of the normal curve requires an interval of $\log \sigma_r / 0.4343 = 0.23026$ on the X scale. It should be carefully noted that this equation merely gives the position on the X scale at which

the normal z ordinates are plotted, and not the position which they take when transformed to the skewed curve, as explained below.

TABLE I

NORMAL CURVE (z ON t OR X_N) AND RELATED LOGARITHMIC NORMAL CURVE (Y ON X) FOR $\log \sigma_r = 0.1$

(Cf. equations 3, 5, and 7)

(1) x/σ, or t	(2) X_N	(3) z	(4) $\log X$	(5) X	(6) Y
−3.5	0.194	0.00087	−0.35	.447	0.002
−3.0	0.309	0.00443	−0.30	.501	0.009
−2.5	0.424	0.01753	−0.25	.562	0.031
−2.0	0.539	0.05399	−0.20	.631	0.086
−1.5	0.655	9.12952	−0.15	.708	0.183
−1.0	0.770	0.24197	−0.10	.794	0.305
−0.5	0.885	0.35207	−0.05	.891	0.395
0.0	1.000	0.39894	0.00	1.000	0.399
0.5	1.115	0.35207	0.05	1.122	0.314
1.0	1.230	0.24197	0.10	1.259	0.192
1.5	1.345	0.12952	0.15	1.413	0.092
2.0	1.461	0.05399	0.20	1.585	0.034
2.5	1.576	0.01753	0.25	1.778	0.010
3.0	1.691	0.00443	0.30	1.995	0.002
3.5	1.806	0.00087	0.35	2.239	0.000

The relation of the normal (z) to the skewed (Y) curve may be informally visualized in terms of the normal and skewed binomials previously discussed. If the normal curve is thought of as made up of very narrow frequencies, or ordinates having a minimum width, the skewed curve may similarly be thought of as made up of these same ordinates suitably adjusted. To the left of $X=1$ these ordinates are narrowed as compared with the normal to fit the progressively narrowing class intervals on the X scale, and heightened to preserve the original area. To the right of $X=1$ the class intervals progressively increase, and the frequencies correspondingly decrease in height. At the geometric mean of Y, i.e, at $X=1$, z and Y coincide. These relationships arise from the fact that the original t scale proceeds by arithmetic increments, while the same points transferred to the X scale proceed by exponential increments, though the arbitrary arithmetic scaling of X obscures this fact.[4] From the nature of the scaling it follows that the position of any transformed z ordinate on the X scale is

[4] The standard deviation ratio takes the place of the constant, C, in the skewed binomial previously discussed, as is apparent in equation 4. It may also be noted that the normal scale, t, is plotted on the X scale in terms of natural logarithms, as is indicated by the use of the modulus, 0.4343, in equation 3 as written, as well as in the form

$$t = \frac{0.4343(X_N - 1)}{\log \sigma_r}.$$

$$X = \sigma_r^t \tag{4}$$

or,

$$\log X = t \log \sigma_r \tag{5}$$

and

$$t = \log X \div \log \sigma_r. \tag{6}$$

And the height of the same ordinate is

$$Y = z/X. \tag{7}$$

With these relationships at hand, it is easy to compute from an ordinary table of the normal curve all the items required for the chart, as illustrated in Table I.

It might be assumed without further proof that the area between any two given values of t on the normal curve would equal the area between the corresponding X values included under the skewed curve Y. This would appear to be obvious in as much as the transformation of the width of the ordinates is offset by the change in the height. And, indeed, it is a principle of the calculus that integrals related by such transformations as are here assumed are equal.

For the purpose of determining areas, or probabilities, between given points on the logarithmic skewed curve, as between the mean and $\pm 2\sigma$ or $\pm 3\sigma$, attention must next be given to the determination of moments of that curve. The general moment about the origin, ν_m, where the geometric mean is unity, may be shown to be related to $\log \sigma_r$ by

$$\nu_m = a^{m^2/2}, \tag{8}$$

where

$$\log a = (\log \sigma_r)^2/0.4343, \tag{9}$$

and m indicates the moment order. This relation may be shown as suggested in a previous reference (this JOURNAL, December 1925, pp. 467 ff.) or by a process of integration.

Thus, M, the arithmetic mean, is

$$M = \nu_1 = a^{1/2} \tag{10}$$

and the second moment about the origin, ν_2, is

$$\nu_2 = a^2 \tag{11}$$

while the variance is

$$\sigma^2 = a^2 - a = a\,(a - 1), \tag{12}$$

since $\sigma^2 = \nu_2 - \nu_1^2$.

If the skewness of a population and of its sample means is to be measured by β_1, this measure for both the population (x) and the sample means ($\bar{x}$) may readily be expressed as follows:

$$\beta_{1:x} = a^3 + 3a^2 - 4 \tag{13}$$

and

$$\beta_{1:\bar{x}} = \beta_{1:x} \div N. \tag{14}$$

If β_1 is known, σ_r may obviously be discovered by the relation noted above, namely, $\log a = (\log \sigma_r)^2/0.4343$.

In view of the preceding considerations, the areas under the logarithmically skewed curve indicating the usual levels of significance may be calculated. Somewhat abbreviated in respect to decimals carried the calculations take the following form:

If $\log \sigma_r = 0.1$ is assumed, then $\log a = (\log \sigma_r)^2/0.43429 = 0.0230259$, and $a = 1.05445$; $M = \sqrt{a} = 1.02686$; $\sigma = \sqrt{a(a-1)} = 0.23961$; and the upper limit of the 0.99 probability as usually measured is $2.57583\sigma = 0.61720$ above the mean on the X scale, or $X = 1.64407$. The corresponding t, or x/σ, may then be found as $\log X \div \log \sigma_r$, and the area read from a table of the normal curve. The entire process designed to determine the required areas, or probabilities, may be summarized as follows:

Skewed curve	(1) X	(2) $\log X$	(3) (2) $\div \log \sigma_r$	(4) Area (cf. table)	(5) Area from M
$M - 2.576\sigma$	0.40966	−0.38758	−3.87577	0.49995	
					0.54578
M	1.02686	0.01151	0.11513	0.04583	
					0.43875
$M + 2.576\sigma$	1.64407	0.21592	2.15920	0.48458	
					0.98453

In the case thus summarized, the total area or probability computed from the slightly skewed theoretical curve between the limits $M \pm 2.576\sigma$ is found to be 0.9845 instead of 0.99 as usually assumed when the effect of skewness is neglected. Though this error of interpretation is only 0.6 per cent high in the aggregate, it is 9 per cent low on the negative side of the mean, and 13 per cent high on the positive side of the mean.

The results of similar calculations for various values of the logarithmic standard deviation at three different levels of significance are set forth in Table II. Particularly in respect to areas below and above the mean, the results substantiate what one would be led to expect from observation, namely, that the error involved in applying the probabilities of the normal curve to a skewed curve of the type here discussed generally increases with the degree of skewness. Apparently the total error of the 0.50, 0.95 and 0.99 probabilities may usually be neglected for practical purposes if log σ_r is less than 0.20, or Sk_Q less than 0.15. But errors involving the setting of fiducial limits in logarithmic skewed curves are more serious, and may well be avoided by reducing in effect such distributions to the normal type by computations based on log X instead of X, or by other suitable corrective methods.

TABLE II

AREAS IN LOGARITHMIC NORMAL CURVES OF SPECIFIED LOGARITHMIC STANDARD DEVIATIONS

log σ_r	Parameters			Area, $M \pm 0.6745\sigma$		
	Sk_Q	M	σ	Below M	Above M	Total
0.01	0.0078	1.0003	0.0231	0.2532	0.2481	0.5013
0.10	0.0775	1.0269	0.2396	0.2810	0.2275	0.5085
0.20	0.1541	1.1119	0.5404	0.3255	0.2101	0.5376
0.30	0.2288	1.2695	0.9927	0.4054	0.1960	0.6014
0.40	0.3010	1.5283	1.7663	0.5584	0.1839	0.7423
0.50	0.3699	1.9401	3.2254	0.7176	0.1729	0.8904
1.00	0.6507	14.1675	200.2169	0.8752	0.1082	0.9834
	Area, $M \pm 1.960\sigma$			Area, $M \pm 2.576\sigma$		
log σ_r	Below M	Above M	Total	Below M	Above M	Total
0.01	0.4815	0.4685	0.9501	0.5005	0.4896	0.9901
0.10	0.5403	0.4142	0.9545	0.5458	0.4388	0.9845
0.20	0.5891	0.3648	0.9539	0.5891	0.3878	0.9769
0.30	0.6351	0.3194	0.9546	0.6351	0.3389	0.9740
0.40	0.6774	0.2821	0.9595	0.6774	0.2975	0.9750
0.50	0.7176	0.2491	0.9667	0.7176	0.2608	0.9784
1.00	0.8752	0.1203	0.9955	0.8752	0.1216	0.9968

LINEAR REGRESSION WHEN THE STANDARD DEVIATIONS OF ARRAYS ARE NOT ALL EQUAL

By G. A. Baker
University of California at Davis

Many textbooks on statistics consider regression only for the case when the standard deviations of all arrays are the same. The implication is that the formulas given are adequate for the cases met in practice and that, even though the standard deviations are not strictly equal, the usual formulas are still valid. Consequently many applications have been made to date without much attention to the equality of the standard deviations of the arrays.[1]

It has been recognized since the days of Gauss that the standard deviations may vary from point to point. The methods for handling varying standard deviations are found in many writings on least squares and are used by some.[2] Nevertheless, the ignorance of varying standard deviations is very widespread as well as the indifference to their importance.

In any set of two-variable data in which the mean of one variable increases from zero or a small value to a large value for values of the other variable there is opportunity for the spread of the arrays to increase as the means of the arrays increase. An excellent example of this situation has been given by Dr. Brady[3] who has shown that the standard deviations of the amounts expended on different phases of living have increased with income. Dr. S. C. Salmon has recently forcibly called attention[4] to the increase of the standard deviations of agricultural yields with the increase in yield. Anything that grows, as a peach embryo, say, must start with size zero and zero standard deviation. As the size increases, the opportunity for variation comes into existence. Thus, peach embryos measured at different times show different standard deviations. In view of the many cases of practical importance where the standard deviations change from array to array, it becomes very important to consider regression freed of the assumption that the standard deviations of arrays are equal.

[1] G. Udny Yule, *An Introduction to the Theory of Statistics*, C. Giffin & Co., Ltd., London, Tenth Edition (1932), p. 177 et seq.

[2] A classic paper on this subject is one by Charles H. Kummell in *The Analyst* (Des Moines), 1879, pp. 97–105. A recent paper that really contains something new is by Arthur J. Kavanagh in *The Annals of Mathematical Statistics*, Vol. 12, No. 1, March 1941, pp. 111–114.

[3] "Variations in Family Living Expenditures," this Journal, Vol. 33, No. 202 (1938), pp. 385–389.

[4] "Generalized Standard Errors for Evaluating Bunt Experiments with Wheat," *Journal American Society of Agronomy*, Vol. 30, No. 8 (1938), pp. 647–663.

The purpose of this paper is to present a method for obtaining the "best" least-square line of regression when the standard deviations of arrays are not equal. No contribution is made to the theory of least-squares since the method is simply an application of the extended theorem of Markoff on least squares as given by F. N. David and J. Neyman.[5] David and Neyman do not claim that their results are new but do claim that their proofs are simpler and the results are more easily applied when expressed in their notation. Their notation and methods are used here merely as a matter of convenience not as a matter of necessity. The method is applied in detail to rat data supplied by Dr. H. H. Cole of the Animal Science Division.

All of the conditions for the straightforward application of the extended theorem of Markoff to the case of linear regression with varying standard deviations are satisfied except

$$\sigma_i^2 = \frac{\sigma^2}{P_i} \tag{1}$$

where P_i's are known positive numbers. The derivation for the case of unknown P_i's and linear regression follows through step-by-step the derivation given by David and Neyman for the case of linear regression and equal standard deviations except that the P_i's must be kept in. Later, in order to apply the results it is necessary to make a choice for the P_i's.

If we do this we obtain

$$F(X) = \frac{\begin{vmatrix} \sum y_iP_i & \sum x_iy_iP_i \\ \sum x_iP_i & \sum x_i^2P_i \end{vmatrix} - X \begin{vmatrix} \sum y_iP_i & \sum x_iy_iP_i \\ \sum P_i & \sum x_iP_i \end{vmatrix}}{\begin{vmatrix} \sum P_i & \sum x_iP_i \\ \sum x_iP_i & \sum x_i^2P_i \end{vmatrix}} \tag{2}$$

as the equation of the "best" line and

$$\mu_P^2 = -\frac{\begin{vmatrix} \sum y_i^2P_i & \sum y_iP_i & \sum x_iy_iP_i \\ \sum y_iP_i & \sum P_i & \sum x_iP_i \\ \sum x_iy_iP_i & \sum x_iP_i & \sum x_i^2P_i \end{vmatrix} \begin{vmatrix} 0 & 1 & X \\ 1 & \sum P_i & \sum x_iP_i \\ X & \sum x_iP_i & \sum x_i^2P_i \end{vmatrix}}{(n-2)\begin{vmatrix} \sum P_i & \sum x_iP_i \\ \sum x_iP_i & \sum x_i^2P_i \end{vmatrix}^2} \tag{3}$$

[5] "Extension of the Markoff Theorem on Least-Squares," *Statistical Research Memoirs*, Vol. II (1938), pp. 105–116.

as the variance of the ordinate of the "best" line. In (2) and (3) X is the running abscissa of the fitted line, x_i and y_i are the observed abscissas and ordinates, P_i's are unknown weights that are to be determined and n is the number of pairs of values of x and y. If all the P_i's are equal, these equations reduce to the usual case treated in most textbooks.

If the x_i's are equal in subgroups then the P_i's in (2) and (3) can be estimated. If the x_i's are not equal in subgroups then (2) and (3) are indeterminate unless some information outside of the sample is available. It should be noted that any numbers proportional to the P_i's will serve as well as the P_i's themselves.

As an example the method of fitting linear regression lines to data in which the standard deviation varies from array to array is applied to some data on the size of ovary of rats that have been stimulated by the injection of a hormone. These data have been partially published by Cole and Saunders[6] and have been furnished in detail by Dr. Cole. The data are given in Table I. The array distributions show no evidence of non-normality. The means of ovarian weights plotted against the logarithms of the dosage in rat units is apparently linear. The standard deviations of arrays plotted against the dosage in rat units is approximately linear.

The standard deviations were smoothed by fitting a least square line giving equal weight to each, obtaining

$$y = -4.06 + 1.85x \tag{4}$$

where y is standard deviation and x is dosage in rat units. As a second smoothing (2) was used with the P_i's taken as inversely proportional to the variances of each array as determined from the first line. The P_i's used were 10.34, 2.33, 1.00, 0.35, 0.18, obtained by putting $x=4$, 6, 8, 12, 16 successively in (4) and then dividing each y by the y for $x=8$. The necessity of weighting the observed standard deviations with P's that are inversely proportional to the smoothed variances of each array in smoothing the variances has been almost universally overlooked but is of considerable importance. The second smoothing of the standard deviations gave

$$y = -3.29 + 1.88x. \tag{5}$$

This process could be continued further.

[6] "The Concentration of Gonad-stimulating Hormone in Blood Serum and of Oestrin in the Urine Throughout Pregnancy in the Mare," *Endocrinology*, Vol. 19 (1935), pp. 199–208.

TABLE I

OVARIAN WEIGHTS IN MILLIGRAMS OF 96 RATS RECEIVING VARYING DOSAGES OF SERUM VII

Number	Dosage in rat units				
	4	6	8	12	16
1	26	31	44	97	29
2	18	48	47	63	51
3	23	40	52	43	102
4	21	43	46	36	75
5	23	26	32	65	64
6	26	56	38	71	47
7	40	35	39	55	80
8	29	31	46	101	79
9	28	37	37	62	57
10	35	30	57	70	45
11	26	30	42	50	76
12	26	39	34	83	112
13	28	40	34	73	136
14	30	34	52	70	88
15	28	49	56	50	81
16	25	44	39	46	44
Mean	27.00	38.31	43.44	65.25	72.88
Variance	25.38	63.28	58.28	303.32	735.26

Hence the P_i's for use in fitting the regression line to the means of the ovary weights are proportional to

$$\begin{aligned} P_1 &= 7.72 \qquad P_3 = 1.00 \qquad P_5 = .19. \\ P_2 &= 2.16 \qquad P_4 = .37 \end{aligned} \tag{6}$$

The P's are obtained by putting $x=4$, 6, 8, 12, 16, successively in (5) and then dividing each y by 15.04, the value of y for $x=8$.

The regression line of ovarian weight, y, on logarithm to base 10 of dosage in rat unit, x, is

$$y = -10.15 + 62.37x. \tag{7}$$

For comparison, the line fitted by giving equal weights to the mean of each array is

$$y = -22.91 + 79.14x. \tag{8}$$

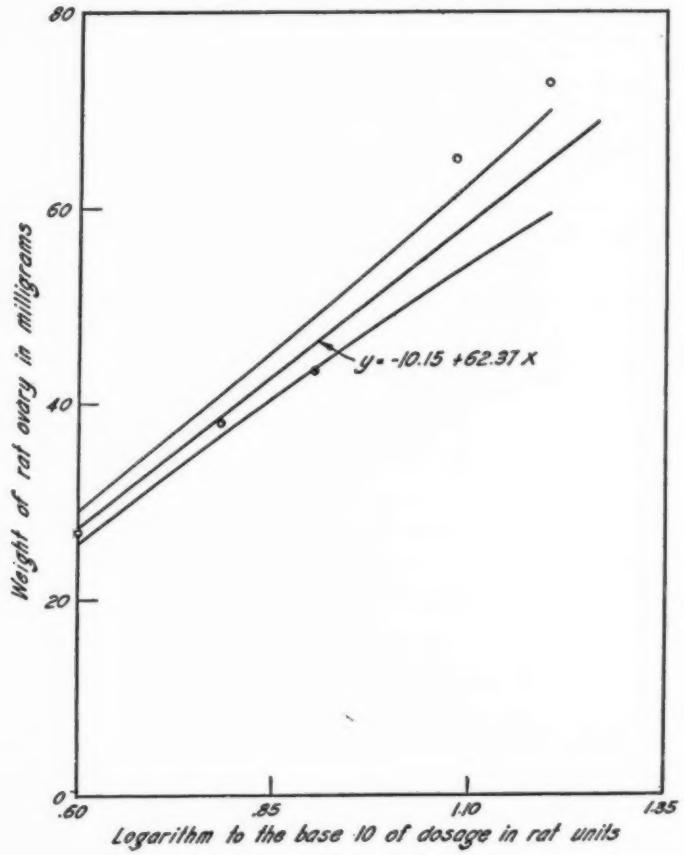
The variances of the ordinates of (7) from (3) are given by

$$\mu_P^2 = 93.92x^2 - 128.73x + 46.22 \tag{9}$$

CHART I

OBSERVED MEAN OVARIAN WEIGHTS COMPARED WITH THE LINEAR REGRESSION LINE INFLUENCED BY THE SPREAD OF THE DATA

The curved lines indicate the standard deviations of the corresponding ordinates. These estimates of the standard deviations of the ordinates are also influenced by the spread of the data.



and of (8) from the usual formula by

$$\mu_P^2 = 69.22x^2 - 126.45x + 60.89. \tag{10}$$

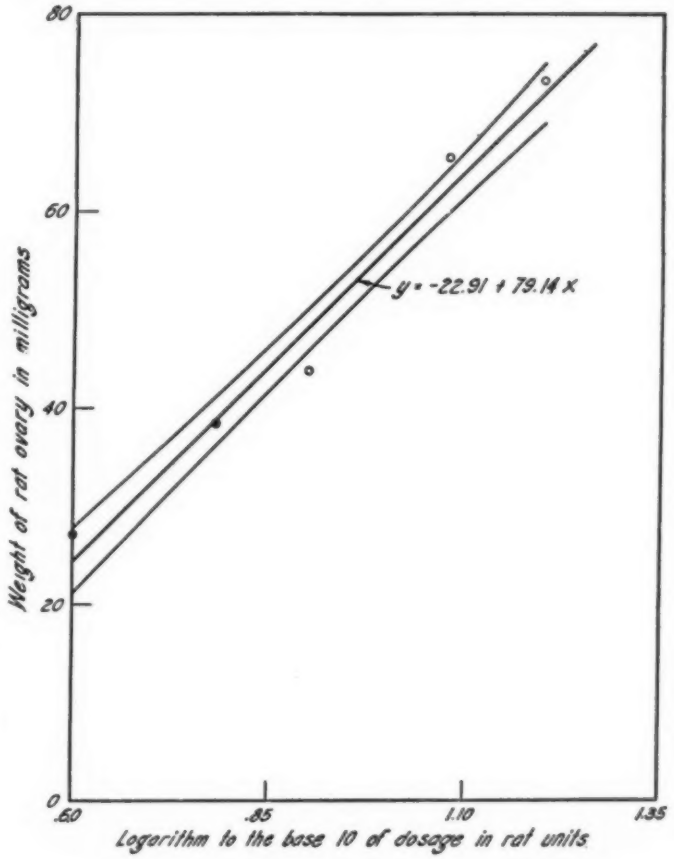
The line (7) and the observed values of the means of groups of 16 are shown in Chart I. The standard deviation of each ordinate as given by (9) is plotted on either side of the line.

The line (8) and the observed values are shown in Chart II, along with the standard deviation of the line as given by (10).

CHART II

OBSERVED MEAN OVARIAN WEIGHTS COMPARED WITH THE LINEAR REGRESSION LINE FITTED BY GIVING EQUAL WEIGHT TO EACH MEAN

The curved lines indicate the standard deviations of corresponding ordinates computed by assuming the spread of the data is uniform for all dosages of hormone.



The two pictures are quite different. The constants of the lines are greatly changed by proper fitting. The standard deviations of the ordinates for given dosages are very erroneously indicated by the usual formula. If it is important to obtain the "best" line the formulas and methods of this paper should be followed whenever it is possible to do so.

This particular example has certain characteristics that, perhaps, should be discussed further. The ordinary method gives a line (Chart

II) which fits quite well at either end and which shows no evidence of systematic deviation from linearity. This fit was exhibited to several "practical" statisticians and pronounced by them to be as good as could reasonably be expected and was accepted as perfectly satisfactory. No question of the linearity of the regression was raised.

When the straight line was properly fitted (Chart I) the data began to show a systematic deviation from linearity as the dosage was increased even though the logarithmic transformation was used. This fact suggests that the regression is really not linear. Thus, an incorrect method of fitting gave no noticeable indication of non-linearity of regression while correct fitting immediately suggested that linearity of regression might well be questioned in this particular instance. The example is retained because it shows two distinct things very well. First, the example shows the great effect that considering the varying standard deviations has on the fitted regression line. Second, the example shows that incorrect fitting may obscure the real nature of the regression.

It may be noted that when the x_i's are fixed and the corresponding arrays are normal but with different variances, if the equation of the regression line is written as

$$y = a + bx, \tag{11}$$

that the estimates of a are normally distributed with variance

$$\sigma_a^2 = \frac{\sum \sigma_i^2}{n^2} \tag{12}$$

and that the estimates of b are normally distributed with variance

$$\sigma_b^2 = \frac{\sum (x_i - \bar{x})^2 \sigma_i^2}{[\sum (x_i - \bar{x})^2]^2} . \tag{13}$$

The correlation coefficient between estimates of a and b is

$$r_{ab} = \frac{\sum (x_i - \bar{x})\sigma_i^2}{\sqrt{\sum \sigma_i^2}\sqrt{\sum (x_i - \bar{x})^2 \sigma_i^2}} . \tag{14}$$

The joint distribution of estimates of a and b is a normal correlation surface determined by (12), (13), and (14).

ACCUMULATING CUBES WITH PUNCH CARDS

By Marion M. Sandomire[1]

A method for obtaining the sum of squares of a series of numbers by means of punch cards was set forth by A. E. Brandt.[2] This has been elaborated upon in a number of other papers.[3]

Under certain conditions the sum of cubes might be required. If the original numbers and their squares have been included on the cards, present methods of successive digiting will yield the sum of cubes. If squares have not been punched, a multiplying punch machine will yield the results ultimately. However, if the card has been filled, and a multiplying punch is not available, another method is needed.

Other procedures have, undoubtedly, been used. In the case of two-digit numbers, it is easy to sort the cards so that a frequency distribution of the numbers from 1 to 99 is obtained. Then, with the assistance of a table of cubes, a maximum of 99 multiplications of the cubes and their corresponding card count yields the desired result. But, this process is somewhat lengthy both as to sorting and multiplying. Improved variations, as suggested by Dwyer,[4] include (1) the calculation of the values of xf_x on the tabulator and multiplication by the squares of the x values, (2) the cumulations of the values of xf_x and the multiplication of these cumulations by the successive odd integers, and (3), if a summary punch is available, the use of successive cumulations to reduce the number of multiplications to three. It is the present purpose, however, to show how the sum of the cubes can be obtained by the simple method of digiting.

CUBING TWO-DIGIT NUMBERS

The digit in the ten's place will be called b and in the unit position a. Then first, sorting on b, tabulate (controlling on b) the subtotals of b and a. Second, sorting on a, tabulate (controlling on a) the subtotals of b and a. The tabulation sheet appears as Figure I.

[1] It is a pleasure to acknowledge the suggestions made by Dr. Paul S. Dwyer and Dr. A. E. Brandt for improving the presentation of this method.

[2] A. E. Brandt, "The Use of Machine Factoring in Multiple Correlation," this Journal, Vol. 23 (1928), pp. 291–295.

[3] Richard Warren and Robert M. Mendenhall, "Mendenhall-Hollerith Correlation Method," Columbia University Statistical Bureau Document No. 1, New York, 1929.

G. W. Baehne, editor, "Practical Applications of the Punched Card Method in Colleges and Universities," Columbia University Press, New York, 1935.

J. L. Comrie, G. B. Hey, and H. G. Hudson, "Application of Hollerith Equipment to an Agricultural Investigation," Supplement to the *Journal of the Royal Statistical Society* Vol. 4 (1937), No. 2, pp. 210–224.

[4] Paul S. Dwyer, "The Computation of Moments with the Use of Cumulative Totals," *Annals of Mathematical Statistics*, Vol. 9 (1938), pp. 288–304.

FIGURE I

Control	Subtotals	
b	b	a
9	126	51
8	40	25
7	77	52
6	54	51
5	30	23
4	48	60
3	33	53
2	32	72
1	11	46
a	b	a
9	54	99
8	31	72
7	60	91
6	25	30
5	63	75
4	37	36
3	17	9
2	83	30
1	49	11

In a calculating machine accumulate[5] cross products of

(I) square of control b and subtotal of b
(II) square of control b and subtotal of a (A)
(III) square of control a and subtotal of b
(IV) square of control a and subtotal of a.

The sum of the cubes will then equal

$$1{,}000(\text{I}) + 300(\text{II}) + 30(\text{III}) + (\text{IV}). \qquad \text{(B)}$$

The proof of this lies, simply, in the fact that any number has been put in the form $10b+a$, and its cube,

$$(10b + a)^3 = 10^3b^3 + 3 \cdot 10^2b^2a + 3 \cdot 10ba^2 + a^3.$$

The sums take care of all values.

In the illustration shown, 100 cards were used. If the values are not too large, sums (I) and (II) can be obtained in the calculating machine in the same operation; similarly, for sums (III) and (IV). Carrying out the indicated accumulations, the four sums in (A) are found to be:

(I) 20,437
(II) 12,461
(III) 12,899
(IV) 20,829.

Substitution in (B) yields the final result of 24,583,099.

[5] Whenever an accumulation involving powers of the control values is mentioned, it should be realized that the numbers from 1 to 9, only, are involved.

This process requires a total of 36 plus 4 multiplications. The final 4 may be avoided by combining the factors, that is, the squares of the controls and the binomial coefficients, as given in Table I.

TABLE I

FACTORS OF SUBTOTALS FOR CALCULATING FOUR SUMS AS DEFINED IN (A)

Control value	Sum			
	(I)	(II)	(III)	(IV)
9	81,000	24,300	2,430	81
8	64,000	19,200	1,920	64
7	49,000	14,700	1,470	49
6	36,000	10,800	1,080	36
5	25,000	7,500	750	25
4	16,000	4,800	480	16
3	9,000	2,700	270	9
2	4,000	1,200	120	4
1	1,000	300	30	1

If the subtotals used in calculating (I) and (IV) are not too large, they can be placed in the bed of the calculating machine at the same time, the first set three places to the left, and the factors for sum IV, only, used. This reduces the number of multiplications to 27. This short cut is limited by the size of the calculating machine. It would be possible to combine subtotals for (II) and (III) thus reducing the number of multiplications to 18. But the mental arithmetic due to overlapping would not be warranted.

The same result may be obtained with the use of progressive totals. In this case, control 9 must be printed first, and so on in reverse order. Then each of the four sums required above can be calculated by adding the products of the nine progressive totals and the factors:

17, 15, 13, 11, 9, 7, 5, 3, and 1, respectively.

This is a consequence of the fact that the sum of the first n odd integers is equal to the square of n. Use of the smaller factors with this method is a timesaver.

The final result is obtained as in expression (B). A table of factors corresponding to the first may be made up. The same final short cuts should be used when feasible.

The value in using the factors in either of the tables to satisfy formula (B) depends upon the computer. Since their use obviates recording, by hand, sums (I), (II), (III), and (IV), a possibility for error is avoided.

TABLE II

FACTORS OF PROGRESSIVE TOTALS FOR CALCULATING FOUR SUMS AS DEFINED IN (A)

Control value	Sum			
	(I)	(II)	(III)	(IV)
9	17,000	5,100	510	17
8	15,000	4,500	450	15
7	13,000	3,900	390	13
6	11,000	3,300	330	11
5	9,000	2,700	270	9
4	7,000	2,100	210	7
3	5,000	1,500	150	5
2	3,000	900	90	3
1	1,000	300	30	1

CUBING THREE-OR-MORE DIGIT NUMBERS

This process may be expanded to numbers of more than two digits, but not all arrangements require fewer operations than the frequency-distribution method. This latter method is, again, subject to the same type of improvements as stated above. It is the purpose of the remainder of the paper to outline a digiting method for obtaining the sum of the cubes of a series of three-digit numbers. The difficulty introduced by a term involving the product of the three digits, such as Σabc, must be surmounted.

The cube of a three-digit number may be represented as follows:

$$(100c + 10b + a)^3 \equiv (10^2c + 10b + a)(10^2c + 10b + a)^2.$$

This identity is given so that we may break up the first factor. Obviously, each of its three terms will be a multiplier of the same squared quantity. But, the square will be represented in a different form in each case. When squaring, combine two of the terms to create a binomial. The term remaining singly should correspond to the multiplier. We have then for the cube:

$$\begin{aligned} &10^2c[10^4c^2 + 2\cdot 10^2c(10b + a) + (10b + a)^2] \\ &\quad + 10b[10^2b^2 + 2\cdot 10b(10^2c + a) + (10^2c + a)^2] \\ &\quad + a[a^2 + 2\cdot a(10^2c + 10b) + (10^2c + 10b)^2]. \end{aligned}$$

Let the frequency be represented by n whenever needed. Summing the above expression for all values, the total sum of cubes is

$$
\begin{aligned}
&10^6\sum_{i=9}^{1} c_i^3 n_i + 2\cdot 10^4\sum_{i=9}^{1}\left[c_i^2\sum(10b+a)_i\right] + 10^2\sum_{i=9}^{1}\left[c_i\sum(10b+a)_i^2\right] \\
&\quad + 10^3\sum_{j=9}^{1} b_j^3 n_j + 2\cdot 10^2\sum_{j=9}^{1}\left[b_j^2\sum(100c+a)_j\right] \\
&\quad + 10\sum_{j=9}^{1}\left[b_j\sum(100c+a)_j^2\right] + \sum_{k=9}^{1} a_k^3 n_k \\
&\quad + 2\sum_{k=9}^{1}\left[a_k^2\sum(100c+10b)_k\right] + \sum_{k=9}^{1}\left[a_k\sum(100c+10b)_k^2\right].
\end{aligned}
\tag{C}
$$

The tabulation consists of progressive totals to be assigned to each value of c, b, and a, in turn, of the other two columns together. Thus, the first main sort is made on c, and for each value of c, a sort is made first on b and then on a to obtain progressive totals of $(10b+a)$.

The second main sort is made on b. This process automatically retains the c arrangement, but a secondary[6] c sort is needed, nevertheless, to facilitate insertion of blank cards for missing values of c to get correct progressive totals of $(100c+a)$.

The third main sort is the a sort. Again, in general, the secondary sorts on c and b are needed in the process of getting progressive totals of $(100c+10b)$.

The tabulation sheet will appear somewhat as Figure II. Extra space is allowed for tabulating or writing in sums of squares later. The card count is required for each of the 27 groups.

The 27 sets of 18 progressive totals are added to obtain sums of squares. The final result is obtained by $27\times3=81$ multiplications, plus 9 to allow for factors. Or this last can be done in the calculating machine without recording by first obtaining the sum of the three center terms in (C), which involve square of control, doubling the result, and then adding in all other terms, taking care of the factors at the same time.

At this point it might be well to note that the card count as shown must be corrected for the number of blank cards inserted in obtaining the progressive totals. This is easily picked up according to the number of blank or unprinted spaces in the control column, and a check is available for each group by comparison of the two runs made. The corrected card count is shown in parenthesis.

No trouble is apt to be experienced in obtaining the sum of squares of $(10b+a)$. But, caution should be exercised when obtaining those of

[6] The term "secondary" will be used in connection with the sorting of a group of cards applying to a particular value of c, b, or a.

FIGURE II

Part A

c Sort	Control	Card count	Progressive totals of (100c+a)	Sums of squares of (10b+a)	b Sort	Control	Card count	Progressive totals of (100c+a)	Sums of squares of (100c+a)	a Sort	Control	Card count	Progressive totals of (100c+10b)	Sums of squares of (100c+10b)
900	70		152*		90	800		808*		9	900		3730*	
	60		221*			700		2211*					3730*	
	50		388*			600		3421*					3730*	
	40		430*					3421*			600		5720*	
	30		468*			400		4631*					5720*	
	20		566*			300		5240*					5720*	
	10		580*			200		5445*			300		6110*	
	00		589*					5445*					6110*	
		14*	589*				13*(11)	5445*			100		6470*	
												16*(11)	6470*	
	9		166*			9		918*						
	8		282*			8		1726*			90		1080*	
			282*					1726*			80		1760*	
	6		364*					1726*					1760*	
			364*			5		1931*			60		2720*	
	4		452*			4		2335*			50		3820*	
			452*			3		3141*					3820*	
	2		568*			2		3843*					3820*	
	1		589*			1		5145*			20		5360*	
		17*(14)	589*	31,569		0		5445*			10		5470*	
							13*(11)	5445*	3,084,691		00		6470*	
												14*(11)	6470*	5,000,100
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
100	80		243*		10	900		904*		1	900		920*	
			243*					904*			800		1770*	
	60		306*			700		2311*			700		2560*	
	50		422*					2311*			600		3250*	
	40		463*			500		3320*			500		3760*	
			463*			400		4529*			400		5020*	
			463*					4529*			300		5720*	
	10		499*			200		4734*			200		5970*	
	00		509*			100		4950*			100		6390*	
		14*(11)	509*				14*(11)	4950*				14*	6390*	
	9		87*			9		109*			90		1480*	
			87*			8		617*			80		1660*	
	7		161*			7		724*					1660*	
			161*			6		1130*					1660*	
			161*			5		2040*			50		3910*	
			161*			4		2944*			40		4050*	
	3		224*					2944*					4050*	
	2		306*			2		4048*			20		4970*	
	1		429*			1		4950*			10		5890*	
	0		509*					4950*			00		6390*	
		15*(11)	509*	32,797			12*(11)	4950*	2,868,706			17*(14)	6390*	3,829,300

FIGURE II

Part B

c Sort	$10^3[9^3(14)+8^3(5)+\cdots+1^3(11)]$	=20,437,000,000
	$2\cdot10^3[9^3(589)+8^3(273)+\cdots+1^3(509)]$	= 2,773,600,000
	$10^3[9(31,569)+8(19,133)+\cdots+1(32,767)]$	= 149,923,800
		23,360,523,800
b Sort	$10^3[9^3(11)+8^3(9)+\cdots+1^3(11)]$	= 20,829,000
	$2\cdot10^2[9^3(5,445)+8^3(3,135)+\cdots+1^3(4,950)]$	= 260,491,600
	$10[9(3,084,691)+8(1,602,201)+\cdots+1(2,868,706)]$	= 1,265,788,320
		1,547,108,920
a Sort	$9^3(11)+8^3(9)+\cdots+1^3(14)$	= 19,223
	$2[9^3(6,470)+8^3(5,860)+\cdots+1^3(6,390)]$	= 2,988,240
	$9(5,000,100)+8(4,620,300)+\cdots+1(3,829,300)$	= 161,519,800
		164,527,263

$(100c+a)$ and of $(100c+10b)$ to add two ciphers for the c values in each instance, and one cipher for the b values.

Quite some time may be saved if for each main sort the progressive totals for one of the two controls are obtained for all nine digits before obtaining the companion sets with the other control. This avoids the necessity for changing the plugboard for each digit of the main sort.

For the illustrative example used here, the three summations calculated from each main sort are recorded, in order of their appearance in formula (C), at the end of the tabulations. The total of these nine quantities yields the final result, which is found to be 25,072,159,983.

Only 100 cards were used for this illustration since it was desired not to obscure the method with large numerical values. This method, in general, would be of value when 300, or more, items are involved.

Thus a total of 486 operations of addition and 81 of multiplication, or 567 operations as compared with a maximum of 999 multiplications yields the result, and allows for easy verification. It should be noted that, while a maximum of 54 runs through the tabulator is required, only one-ninth of the cards are used during each run.

As pointed out by Dwyer, if the machine is equipped with a digit selector, four class selectors, the usual five x-distributors, and has sufficient adding capacity, it is possible to combine groups of nine of the previous runs into single runs so that but six runs through the tabulator are necessary. This is done with the use of "spread" or "horizontal digiting" in which the values in (C), as for example $(10b+a)$, control-

ling on b, are placed in different columns according to the value of a with the use of the digit selector. It is then possible to obtain all the quantities necessary for the calculation of (C) if a card count for the control digit accompanies the "spread" run. Lack of sufficient adding capacity may necessitate more than six runs. The work can be arranged so that three sorts only are necessary.

For four-digit numbers the sum of the cubes may be obtained in 1,080 calculating-machine operations as compared with a possible maximum of 9,999. The tabulating work would require four main sorts, as determined by the following method of factoring:

$$
\begin{aligned}
&(10^3d + 10^2c + 10b + a)^3 \\
&\quad \equiv (10^3d + 10^2c + 10b + a)(10^3d + 10^2c + 10b + a)^2 \\
&\quad = 10^3d[10^6d^2 + 2\cdot 10^3d(10^2c + 10b + a) + (10^2c + 10b + a)^2] \\
&\qquad + 10^2c[10^4c^2 + 2\cdot 10^2c(10^3d + 10b + a) + (10^3d + 10b + a)^2] \\
&\qquad + 10b[10^2b^2 + 2\cdot 10b(10^3d + 10^2c + a) + (10^3d + 10^2c + a)^2] \\
&\qquad + a[a^2 + 2\cdot a(10^3d + 10^2c + 10b) + (10^3d + 10^2c + 10b)^2].
\end{aligned}
$$

With a maximum of nine digits to be considered for each sort, and progressive totals required for three controls, 108 tabulator runs are indicated. However, since but one-ninth of the cards are used each time, the total number of runs through the tabulator actually is twelve. This is twice the number required for three-digit numbers, or in general $k(k-1)$.

FITTING THE EXPONENTIAL FUNCTION AND THE GOMPERTZ FUNCTION BY THE METHOD OF LEAST SQUARES

By Paul Matthew Stoner
Federal Housing Administration

The general modified exponential function—$y=a+bc^x$—and the logarithm of the Gompertz function—$\log y=\log a+c^x(\log b)$—can be fitted, with any desired degree of accuracy, to a time series by the method of least squares. Not only does the method utilize all the observational data, instead of arbitrarily selected points as used in the usual statistical methods, but it permits the application of significance tests.

In the development of the normal equations and of the technique of calculation, which follows, the logarithmic Gompertz function will be used as an example. However, the outlined method is general and, with the one exception hereafter noted, can be used for either the exponential function or its logarithmic form and for either the function with or without the constant "*a*."

The general formula of the Gompertz curve is $y=ab^{c^x}$; the logarithmic function of which, as stated above, is $\log y=\log a+c^x(\log b)$.

Deriving the three required normal equations by partial differentiation in respect to each of the three unknown constants appearing in the equation, expressing the sum of the squares of the deviations of the observed values from the computed values and equating[1] each of the three partial derivatives to 0, we have:

Normal Equation I.—$\sum\log y=n(\log a)+\log b\ (\sum c^x)$
Normal Equation II.—$\sum c^x(\log y)=\log a\ (\sum c^x)+\log b\ (\sum c^{2x})$
Normal Equation III.—$\sum xc^x(\log y)=\log a\ (\sum xc^x)+\log b\ (\sum xc^{2x})$.

Taking the origin of the argument, x, at 0 and summing the sequences for n years (the last value of the argument being $n-1$), the three normal equations can be written:

$$\text{I.}\qquad \sum_0^{n-1}(\log y)=n(\log a)+\log b\left[\frac{c^n-1}{c-1}\right]$$

$$\text{II.}\qquad \sum_0^{n-1}c^x(\log y)=\log a\left[\frac{c^n-1}{c-1}\right]+\log b\left[\frac{c^{2n}-1}{c^2-1}\right]$$

[1] See exposition of the fundamental theory of the Method of Least Squares in Mansfield Merriman, *Method of Least Squares*, Eighth Edition, Chapter 3.

III.
$$\sum_{0}^{n-1} xc^x(\log y) = \log a\left[\frac{c^n(nc - n - c) + c}{(c-1)^2}\right] + \log b\left[\frac{(n-1)c^{2n+2} - nc^{2n} + c^2}{(c^2-1)^2}\right].$$

Solving for c by simultaneous processing of the above three normal equations and reducing to the simplest form, we have:

$$\left[\frac{(c-1)(c^n-1)\left[\sum_0^{n-1} xc^x(\log y)\right] - \left[(n-1)c^{n+1} - nc^n + c\right]\left[\sum_0^{n-1} c^x(\log y)\right]}{n(c-1)\left[\sum_0^{n-1} c^x(\log y)\right] - (c^n-1)\left[\sum_0^{n-1}(\log y)\right]}\right] = \left[\frac{c(c^{2n}-1) - nc^n(c^2-1)}{n(c^2-1)(c^n+1) - (c+1)^2(c^n-1)}\right].$$

In the solution of this implicit function, the left-hand side will be subsequently referred to as Z_1 and the right-hand side as Z_2.

Since this is an implicit function, the value of c cannot be expressed directly in terms of the known values. It is necessary therefore to resort to finite calculus, successive approximations. Although the method is approximative, any desired degree of accuracy can be obtained. The degree of accuracy can then be positively checked.

Step 1—The first approximation of c is obtained by the method of selected points[2] or preferably by the method of grouped data.[3] By the latter method, for a Gompertz curve, the approximation will probably be accurate to the second decimal place, therefore the result should be rounded off to the nearest hundredth of a decimal. Subsequent approximations will prove whether this rounding off is justified; if it isn't, the further approximations and interpolations will automatically correct the value of c.

Step 2—Substitute the first approximate value of c in the implicit function and solve for Z_1 and Z_2.

Step 3—If Z_1 is greater than Z_2, i.e. the difference between Z_1 and Z_2 (subsequently referred to as W_1) is positive, reduce the value of c by .01 for the second approximation. If W_1 is negative, increase the value of c by .01. This is the second approximation.

Step 4—With the second approximation of c, solve for Z_1, Z_2, and W_2.

Step 5—Continue solving for W's with successive approximations of c (each .01 apart) until the sign of W is reversed from that obtained in the first approximation. However, for four decimal place accuracy in

[2] Davies and Crothers, *Methods of Statistical Analysis*, 1933, Chapter VI.
[3] F. C. Mills, *Statistical Methods*, 1939, pp. 672–673.

the final result at least three approximations should be obtained. As a general rule for increased accuracy, at least as many values of W should be obtained as the degree of accuracy desired less one, i.e. for five place accuracy, four values of W should be computed.

Step 6—Selecting the three derived consecutive values of W, the first of which is the smallest negative value and the other two being the consecutive positive values, construct a difference table to the second difference. For further accuracy, if desired, additional values of W are used and differences of correspondingly higher order are determined.

Step 7—By the application of the Newton-Gregory formula to the method of inverse interpolation by successive approximations[4] the value of c is derived. By using a difference table of the second order, the interpolated value of c, as previously stated, will ordinarily be accurate to four decimal places.

Step 8—For a positive check of fourth place accuracy, compute the two values of W for the derived value of $c \pm 5\ (0.1)^5$. If the value of W in one case is positive and in the other negative it follows that the true value of c (which would be when $W = 0$) lies between $c \pm 5\ (0.1)^5$. Hence it can be absolutely inferred that the true value of c has been exactly rounded off to four place accuracy. If this positive check does not hold, it is necessary to carry the Newton-Gregory formula to a difference of a higher order.

If accuracy of higher order is being checked, limits of correspondingly higher degree are used.

Step 9—Substitute the derived value of c in the normal equations and solve simultaneously for the other two unknown constants a and b.

Final Check:

1. The sum of the deviations of the computed values of log y from the log of the observed values, allowing for possible error due to rounding off, must equal 0.

2. The sum of the squares of the deviations should be less than the sum of the squares of the deviations for any other values of the constants a, b, and c.

In applying the above outlined technique to exponential functions other than the Gompertz, one variation must be introduced. The Gompertz function as it is generally used has a value of c that is less than 1. However, when the value of c is more than 1, the procedure outlined in Step 3 is reversed. In such instances, if after substituting

[4] H. Freeman, *Actuarial Mathematics*, 1931, Chapter VI.

the derived first approximation of c, a value for W is secured which is positive, the successive approximation of c is increased, and vice versa.

The method outlined above is also applicable for the negative value of c, although the nature of the curve is entirely different from the exponential function or from the Gompertz curve with positive c values, being either a converging or diverging oscillating curve.

No difference in the basic procedure is required for fitting a function of negative c value; the sign of the c being tentatively established by the first approximation in Step 1. This negative value is then processed in accordance with the above outlined technique.

The importance of this adaptation of the method of least squares is in its application to the logarithmic fitting of the Gompertz function and to the arithmetic fitting of the common exponential curve. The method, although an approximative one and somewhat tedious, is precise and gives a unique fit by the method of least squares.

FURTHER NOTES ON THE DIFFERENCE BETWEEN INDEX-NUMBER FORMULAS

By Irving H. Siegel
United States Bureau of Labor Statistics

In an earlier paper in this Journal,[1] various expressions were developed for the difference between index numbers computed by use of the Paasche and Laspeyres formulas. In the present paper, consideration will be given to the development of expressions for the difference between any two indexes which are reducible to weighted arithmetic or harmonic means of relatives. Some remarks will also be made concerning the usually unimportant case of the difference between weighted and unweighted arithmetic means.[2]

Since any weighted harmonic mean can also be written as a (differently) weighted arithmetic mean, no loss in generality will result from confinement of the investigation to weighted arithmetic means of relatives. Although the ideal index does not fall within the scope of this paper, its constituents, the Paasche and Laspeyres formulas, do; and so does the important Edgeworth formula, which may be written as a weighted harmonic or arithmetic mean as well as in the familiar aggregative form:

$$\frac{\Sigma(h'+h)g'}{\Sigma(h'+h)g} = \frac{\Sigma(h'+h)g'}{\Sigma\left[\dfrac{(h'+h)g'}{g'/g}\right]} = \frac{\Sigma(h'+h)g\dfrac{g'}{g}}{\Sigma(h'+h)g}. \qquad (1)$$

Let us consider two indexes,

$$I_1 = \frac{\Sigma kX}{\Sigma k} \quad \text{and} \quad I_2 = \frac{\Sigma mX}{\Sigma m},$$

based on the same n relatives; the $x_i = g_i'/g_i$ are the relatives being averaged, and the k_i, m_i are the two sets of weights. Our object is to determine the conditions under which $I_1 \gtreqless I_2$. Now, the difference at any time is

[1] Irving H. Siegel, "The Difference between the Paasche and Laspeyres Index-Number Formulas," this Journal, September 1941, pp. 343–350.

[2] It should be noted that the expressions developed here apply, not only to means of relatives, but to any means of the same general form. Indeed, an illustration given at the end of this paper involves, not the difference between means of relatives for the same time, but between means of index numbers for different points in time.

$$\Delta = \frac{\Sigma kX}{\Sigma k} - \frac{\Sigma mX}{\Sigma m} = \frac{\begin{vmatrix} \Sigma kX & \Sigma k \\ \Sigma mX & \Sigma m \end{vmatrix}}{\Sigma k \Sigma m}. \tag{2}$$

By expansion of the determinant in this expression or from the fact that the matrix of the determinant is the product of two others, i.e.,

$$\begin{Vmatrix} \Sigma kX & \Sigma k \\ \Sigma mX & \Sigma m \end{Vmatrix} = \begin{Vmatrix} k_1 & \cdots & k_n \\ m_1 & \cdots & m_n \end{Vmatrix} \cdot \begin{Vmatrix} X_1 & \cdots & X_n \\ 1 & \cdots & 1 \end{Vmatrix}, \tag{3}$$

we see that the difference is equivalent to

$$\Delta = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \begin{vmatrix} k_i & k_j \\ m_i & m_j \end{vmatrix} \cdot \begin{vmatrix} X_i & X_j \\ 1 & 1 \end{vmatrix}}{\Sigma k \Sigma m}, \tag{4}$$

the numerator of which is a sum of ${}_nC_2$ products. This numerator may also be written in various other ways, since

$$\begin{vmatrix} k_i & k_j \\ m_i & m_j \end{vmatrix} \cdot \begin{vmatrix} X_i & X \\ 1 & 1 \end{vmatrix} = \begin{vmatrix} k_i m_j & k_j m_i \\ 1 & 1 \end{vmatrix} \cdot \begin{vmatrix} X_i & X_j \\ 1 & 1 \end{vmatrix}$$

$$= m_i m_j \begin{vmatrix} \frac{k_i}{m_i} & \frac{k_j}{m_j} \\ 1 & 1 \end{vmatrix} \begin{vmatrix} X_i & X_j \\ 1 & 1 \end{vmatrix}$$

$$= \begin{vmatrix} \frac{k_i}{m_i} & \frac{k_j}{m_j} & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & m_i & 0 & 0 & 0 \\ 0 & 0 & 0 & m_j & 0 & 0 \\ 0 & 0 & 0 & 0 & X_i & X_j \\ 0 & 0 & 0 & 0 & 1 & 1 \end{vmatrix}. \tag{5}$$

In each instance, the double sum has ${}_nC_2$ components.

The sign of the difference between I_1 and I_2 is given by the numerator of (2) or its equivalent (4) and any variant based on (5). In the special case where the rank correlation between the X_i and k_i/m_i is perfect

and positive, $I_1 > I_2$, necessarily; and in the special case where the rank correlation is perfect but inverse, $I_1 < I_2$, necessarily. In the first of these cases, the Pearsonian coefficient of correlation is necessarily positive, and, in the latter, the Pearsonian coefficient is necessarily negative. This is obvious from the fact that

$$\begin{Vmatrix} \Sigma \frac{k}{m} X & \Sigma \frac{k}{m} \\ \Sigma X & n \end{Vmatrix} = \begin{Vmatrix} \frac{k_1}{m_1} & \cdots & \frac{k_n}{m_n} \\ 1 & \cdots & 1 \end{Vmatrix} \cdot \begin{Vmatrix} X_1 & \cdots & X_n \\ 1 & \cdots & 1 \end{Vmatrix}. \tag{6}$$

The matrix on the left-hand side of this identity is the matrix of the determinant giving the sign of the Pearsonian coefficient.

In general, however, the sign of the Pearsonian coefficient is not an accurate indicator of the sign of the difference between I_1 and I_2. The von Bortkiewicz weighted correlation coefficient, on the other hand, provides a trustworthy criterion; its sign is the sign of the difference and, when the difference is zero, it, too, vanishes:

$$\Delta = \frac{\Sigma kX}{\Sigma k} - \frac{\Sigma mX}{\Sigma m} = \frac{\Sigma m \frac{k}{m} X}{\Sigma m \frac{k}{m}} - \frac{\Sigma mX}{\Sigma m}$$

$$= \frac{\Sigma m \frac{k}{m} X - \frac{\Sigma m \frac{k}{m} \cdot \Sigma mX}{\Sigma m}}{\Sigma m \frac{k}{m}}$$

$$= \frac{\Sigma m \left(\frac{k}{m} - \frac{\Sigma m \frac{k}{m}}{\Sigma m} \right) \left(X - \frac{\Sigma mX}{\Sigma m} \right)}{\Sigma m \frac{k}{m}}$$

$$= r_{m:(k/m)\cdot X} \sigma_{m:(k/m)} \sigma_{m:X} \Big/ \frac{\Sigma m \frac{k}{m}}{\Sigma m}. \tag{7}$$

In (7), the only factor which may assume either positive or negative values is $r_{m:(k/m)\cdot X}$, the weighted coefficient of correlation between the k_i/m_i and the X_i. The factors $\sigma_{m:(k/m)}$ and $\sigma_{m:X}$ are the weighted standard deviations of the k_i/m_i and the X_i, respectively. That the sign of the difference would depend on the sign of the weighted correlation coefficient could have been anticipated from the fact that, if the observation equations have the weights m_i, the numerator of (2),

$$\begin{vmatrix} \Sigma kX & \Sigma k \\ \Sigma mX & \Sigma m \end{vmatrix} = \begin{vmatrix} \Sigma m \dfrac{k}{m} X & \Sigma m \dfrac{k}{m} \\ \Sigma mX & \Sigma m \end{vmatrix} = \begin{vmatrix} \Sigma m & \Sigma m \dfrac{k}{m} \\ \Sigma mX & \Sigma m \dfrac{k}{m} X \end{vmatrix},$$

is also the numerator obtained for the weighted regression coefficient in solving by determinants the normal equations for $k/m = a + b_{m:(k/m)\cdot X}X$. The sign of this regression coefficient[3] is, of course, given by $r_{m:(k/m)\cdot X}$.

The expressions already presented and various others that may be developed are analogous to those given in the earlier paper. For the difference between the Paasche and Laspeyres indexes is merely a special case of the difference between two weighted indexes; if, in this special case, the X_i represent price relatives and the m_i represent base-period money values, the k_i/m_i reduce to quantity relatives.

A few comments will now be made on the usually unimportant case of the difference between a weighted and an unweighted average of relatives. In general,

$$\Delta' = \frac{\Sigma mX}{\Sigma m} - \frac{\Sigma X}{n} = \frac{\begin{vmatrix} \Sigma mX & \Sigma m \\ \Sigma X & n \end{vmatrix}}{n\Sigma m} = r_{m\cdot X}\sigma_m\sigma_X \Big/ \frac{\Sigma m}{n}; \qquad (8)$$

that is, the difference may be considered to involve a simple correlation coefficient. The difference may also be considered to involve a double sum; since

$$\left\| \begin{matrix} \Sigma mX & \Sigma m \\ \Sigma X & n \end{matrix} \right\| = \left\| \begin{matrix} m_1 & \cdots & m_n \\ 1 & \cdots & 1 \end{matrix} \right\| \cdot \left\| \begin{matrix} X_1 & \cdots & X_n \\ 1 & \cdots & 1 \end{matrix} \right\|, \qquad (9)$$

[3] It is interesting to note that the weighted correlation and regression coefficients may be written in terms of double sums. The weighted covariance of the k_i/m_i and the X_i is the quotient of the numerator of (4) and $(\Sigma m)^2$. From matrix identities similar to (3), it may be established that the weighted variances are reducible to quotients of double sums of squares of determinants (weighted by m_im_j) and $(\Sigma m)^2$.

Since ${}_nC_3 > n$ for $n > 3$, there is no computational advantage in expressing correlation or regression coefficients in terms of double sums.

Unweighted correlation and regression coefficients may also be expressed in terms of double sums, as may be seen from (6) and analogous matrix identities for unweighted variances.

we may write

$$\Delta' = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \begin{vmatrix} m_i & m_j \\ 1 & 1 \end{vmatrix} \cdot \begin{vmatrix} X_i & X_j \\ 1 & 1 \end{vmatrix}}{n\Sigma m} \cdot \tag{10}$$

The expressions (8) and (10) may be transformed into others analogous to those for the difference between weighted indexes.

Many problems not explicitly involving the difference between such means as are discussed here may nevertheless be reduced to such form. Two illustrations will now be given. Consider, first, a case in which separate data are reported for various grades of the same nominal product. What is the difference between a derived production index in which the several grades are distinguished and weighted and an unweighted aggregative index of the gross quantities? This problem reduces to one of finding the difference between two weighted indexes, for the unweighted aggregative measure is equivalent to a weighted average of relatives:

$$\frac{\Sigma h'}{\Sigma h} = \frac{\Sigma h \dfrac{h'}{h}}{\Sigma h} \cdot$$

The second illustration reduces to the problem of finding the difference between a weighted average and an unweighted one. Consider two series of index numbers—say, for production and employment—on the compound time base $t_1, \cdots, t_n = 100$. Now, the average of the productivity measures computed for the base period from the individual production and employment indexes need not equal 100. Under what conditions will the average of the productivity indexes for this period exceed, equal, or fall short of 100? It should be noted that 100 is also the ratio of the simple average of the production indexes to the simple average of the employment indexes.[4] This ratio may be developed as follows:

$$100 = \frac{\sum_{t=1}^{n} P}{\sum_{t=1}^{n} E} = \frac{\Sigma E \dfrac{P}{E}}{\Sigma E} = \frac{\Sigma \dfrac{P}{E}}{n} + r_{E \cdot (P/E)} \sigma_E \sigma_{(P/E)} \Big/ \frac{\Sigma E}{n} \cdot$$

[4] This ratio is unity rather than 100, but, here as elsewhere, we conform to an innocuous convention and substitute 100.

The simple average of the productivity indexes, on the other hand, is merely $\frac{\sum_{i=1}^{n} \frac{P}{E}}{n}$. Hence, the average of the individual productivity indexes is greater than, equal to, or less than 100—the ratio of the production average to the employment average for the base period—according as $r_{E\cdot(P/E)} \gtreqless 0$.

THE USE OF GROUPED MEASUREMENTS

By Howard L. Jones

In recording and summarizing sample data by class intervals, it is sometimes desirable to obtain the total of the measurements as well as the number of observations for each interval. In recording grouped temperature readings, for example, if 15, 15, 16, and 19 are the observed readings for the class interval 15 to 19, inclusive, it may be desirable to record 65, the total of the readings for that interval, as well as 4, the number of observations.

If it be assumed that the distribution of the data in the universe sampled is continuous and linear for some class interval, the parameters of the line describing the distribution for that interval can readily be estimated by putting

$$n = \int_{p-i/2}^{p+i/2} (a + bx)dx$$

and

$$m = \int_{p-i/2}^{p+i/2} x(a + bx)dx$$

where a and b are the parameters of the fitted line

n = the number of observations for the class interval
m = the sum of the measurements for the interval
i = the width of the interval
and p = the mid-point of the interval.

From these expressions we find

$$n = ai + pbi$$
$$m = np + bi^3/12$$

whence

$$b = 12(m - np)/i^3$$

and

$$a = n/i - pb.$$

If we wish to interpolate within any class interval, we may use

$$n' = i'(a + p'b)$$

and

$$m' = n'p' + bi'^3/12$$

where i' and p' are the width and mid-point, respectively, for the sub-class within the class interval, and n' and m' are the interpolated number of observations and sum of the measurements. A check is obtained by adding the results for all the sub-classes within each class interval and comparing with the actual totals for that interval.

In estimating the standard deviation we may put

$$\sum_{x=p-i/2}^{p+i/2} x^2y = \int_{p-i/2}^{p+i/2} x^2(a+bx)dx \doteq 2mp - np^2 + ni^2/12$$

and

$$\sum_{-\infty}^{+\infty} x^2y \doteq 2\sum mp - \sum np^2 + 1/12\sum ni^2.$$

Higher moments may be estimated in a similar manner.

For data in which x may assume only integral values, i^2-1 may be substituted for i^2.

It may be of interest to consider the error made in assuming that

$$\sum x^2y = 2\sum mp - \sum np^2 + 1/12\sum ni^2.$$

Following the method used by Professor Carver in the *Handbook of Mathematical Statistics* to develop Sheppard's adjustments, we may write

$$n = \int_{-i/2}^{i/2} f(p+h)dh$$

$$m = \int_{-i/2}^{i/2} (p+h)f(p+h)dh$$

$$\sum x^2y = \int_{-i/2}^{i/2} (p+h)^2f(p+h)dh$$

$$f(p+h) = \sum_{r=0}^{\infty} \frac{h^rf^{(r)}(p)}{r!}$$

$$[f^{(k)}(p+h)]_{-i/2}^{i/2} = \sum_{r=0}^{\infty} \frac{i^{2r+1}f^{(k+2r+1)}(p)}{2^{2r}(2r+1)!}$$

$$n = \sum_{r=0}^{\infty} \frac{i^{2r+1}f^{(2r)}(p)}{2^{2r}(2r+1)!}$$

$$m = np + \sum_{r=0}^{\infty} \frac{i^{2r+3}f^{(2r+1)}(p)}{2^{2r+2}(2r+1)!(2r+3)}$$

$$\sum x^2y = np^2 + 2p\sum_{r=0}^{\infty} \frac{i^{2r+3}f^{(2r+1)}(p)}{2^{2r+2}(2r+1)!(2r+3)}$$

$$+\sum_{r=0}^{\infty}\frac{i^{2r+3}f^{(2r)}(p)}{2^{2r+2}(2r)!(2r+3)}$$

$$\sum x^2y = 2mp - np^2 + \sum_{r=0}^{\infty}\frac{i^{2r+3}f^{(2r)}(p)}{2^{2r+2}(2r)!(2r+3)}$$

$$= 2mp - np^2 + \frac{ni^2}{12} + \sum_{r=0}^{\infty}\frac{ri^{2r+3}f^{(2r)}(p)}{2^{2r}3(2r+1)!(2r+3)}$$

$$= 2mp - np^2 + \frac{ni^2}{12} + \sum_{r=0}^{\infty}\frac{i^{2r+5}f^{(2r+2)}(p)}{2^{2r}4!(2r+1)!(2r+3)(2r+5)}.$$

The series indicated by the last term above is the error for a class interval that results from using the abbreviated formula

$$\sum x^2y = 2mp - np^2 + ni^2/12,$$

provided the distribution of the universe is continuous throughout the interval, and m and n are free of sampling errors. By successively subtracting the appropriate multiples of the odd derivatives, $[f^{(k)}(p+h)]_{-i/2}^{i/2}$, from the error term and substituting $r+1$ for r after each subtraction, this series may be written

$$\frac{i^4}{3\cdot 5!}[f'(p+h)]_{-i/2}^{i/2} - \frac{i^6}{3\cdot 7!}[f^{(3)}(p+h)]_{-i/2}^{i/2}$$

$$+\frac{3i^8}{5\cdot 9!}[f^{(5)}(p+h)]_{-i/2}^{i/2} - \frac{5i^{10}}{3\cdot 11!}[f^{(7)}(p+h)]_{-i/2}^{i/2}$$

$$+\frac{691i^{12}}{105\cdot 13!}[f^{(9)}(p+h)]_{-1/2}^{i/2} - \cdots.$$

By summing the results for all the group intervals from $x=a$ to $x=b$ and expressing the coefficients in terms of multiples of the Bernoulli numbers, the error in computing $\sum x^2y$ for this range of x, provided the distribution of the universe is continuous and m and n are free of sampling errors, may be written

$$\frac{B_1B_3i^4}{2}[f'(x)]_a^b - \frac{B_1B_3B_5i^6}{2}[f^{(3)}(x)]_a^b$$

$$+\frac{3B_1B_3B_5B_7i^8}{8}[f^{(5)}(x)]_a^b - \frac{B_1B_3B_5B_7B_9i^{10}}{8}[f^{(7)}(x)]_a^b$$

$$+\frac{B_1B_3B_5B_7B_9B_{11}i^{12}}{80}\left[f^{(9)}(x)\right]_a^b-\cdots$$

where $B_1=1/6$, $B_3=1/30$, $B_5=1/42$, $B_7=1/30$, $B_9=5/66$, and $B_{11}=691/2730$.

TABLE I

BI-HOURLY TEMPERATURE READINGS IN CHICAGO FEBRUARY 26 TO MARCH 4, 1940

Ungrouped Data

Temperature x	Frequency y
15	2
16	1
19	1
21	1
26	1
29	2
30	8
31	8
32	14
33	9
34	15
35	16
36	10
37	6
38	2
$\Sigma y =$	96
$\Sigma xy =$	3,131
$\Sigma x^2y =$	103,893

Grouped Data

Temperature x	Frequency n	Sum of measurements m	Mid-point p	Class interval i	Extensions mp	np^2	ni^2
15–19	4	65	17	5	1,105	1,156	100
20–29	4	105	24.5	10	2,572.5	2,401	400
[illegible]	30	936	31	3	29,916	28,830	270
[illegible]–35	40	1367	34	3	46,478	46,240	360
36–38	18	658	37	3	24,346	24,642	162
	96	3131			103,517.5	103,269	1292

$$\sum np^2 - 1/12\sum n(i^2-1)^* = 103,169.33$$
$$2\sum mp - \sum np^2 + 1/12\sum n(i^2-1)^* = 103,865.67$$

* The substitution of i^2-1 for i^2 in this example was made only for the purpose of obtaining better comparisons between the results of the two methods of computation and 103, 893, which was computed with integral values of x. For ordinary purposes, no such substitution would be made, since the exact temperatures are not restricted to integers.

Table I presents an example showing how the formula

$$\sum x^2y \doteq 2\sum mp - \sum np^2 + 1/12\sum ni^2$$

works with a badly skewed distribution. In this example the true value of $\sum x^2y$ for the sample, except for errors due to rounding off measurements to the nearest whole number, is 103,893. As indicated, the computed value of $\sum x^2y$ based on grouped frequencies with Sheppard's adjustment, is 103,169.33. If the formula based on the sum of the measurements as well as the frequency for each group is used, however, the computed value of $\sum x^2y$ is 103,865.67, the error being less than four per cent of the error when Sheppard's adjustment is used.

A check of the formula

$$\sum x^2y \doteq 2\sum mp - \sum np^2 + 1/12\sum ni^2$$

with theoretical data obtained from tables of the probability curve indicates that this formula is satisfactory for use with grouped normal distributions, the error being of about the same size as when Sheppard's adjustment is made. The formula is also suitable for linear distributions, for which Sheppard's adjustment tends to double the error in the unadjusted $\sum np^2$.

A NEW METHOD FOR OBTAINING THE INVERSE MATRIX

By M. D. Bingham

The inverse matrix is playing an increasingly important role in modern statistical analyses. For example in fitting regression equations by the method of least squares the inverse matrix not only facilitates computation of the regression coefficients but also is required for obtaining the variances and covariances of these coefficients. Again the inverse matrix is essential in discriminant analysis[1] and in testing certain statistical hypotheses concerning sets of means.[2] It is, therefore, important to develop efficient methods for obtaining from a given matrix the numerical values of its inverse.

Aside from certain refinements introduced by the author, the method here presented was developed jointly by Professor Harold Hotelling of Columbia University and Mr. M. A. Girshick of the United States Department of Agriculture. This method has been successfully used in the United States Department of Agriculture, Bureau of Home Economics' Body Measurement Study.[3]

We desire to calculate from a given non-singular, square matrix

$$A = \left\| \begin{matrix} a_{11} \cdots a_{1n} \\ \cdot \\ \cdot \\ \cdot \\ a_{n1} \cdots a_{nn} \end{matrix} \right\| \tag{1}$$

its inverse, say A^{-1}, where by definition[4]

$$A^{-1} = \left\| \begin{matrix} A_{11}/a \cdots A_{n1}/a \\ \cdot \quad \searrow \quad \cdot \\ \cdot \quad \quad \cdot \\ A_{1n}/a \cdots A_{nn}/a \end{matrix} \right\|, \tag{2}$$

A_{ij} being the cofactor of a_{ij} in the determinant a of A.

Consider the characteristic polynomial in λ,

$$\left| A - \lambda I \right| = \lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - \cdots (-1)^n p_n = 0, \tag{3}$$

where I is the identity matrix and $p_r (r = 1, 2, \cdots, n)$ is the sum of the r-rowed principal minors of the matrix A. By a well known theorem in algebra[5] it can be shown that when the matrix A is substituted for λ,

[1] R. A. Fisher, *Statistical Methods for Research Workers*, 7th ed.

[2] Harold Hotelling, "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, Vol. 2 (1931), pp. 360 to 378.

[3] U. S. Department of Agriculture Misc. Publications #365 and #366.

[4] M. Bôcher, *Introduction to Higher Algebra*, pp. 74–75.

[5] *Ibid.*, pp. 282, 296–297.

equation (3) still holds. That is,

$$A^n - p_1A^{n-1} + p_2A^{n-2} - \cdots (-1)^n p_n I = 0 \tag{4}$$

where the symbol 0 represents an n-rowed square matrix whose elements are all zero. Multiplying equation (4) by A^{-1} we get

$$A^{n-1} - p_1A^{n-2} + p_2A^{n-3} - \cdots (-1)^n p_n A^{-1} = 0 \tag{5}$$

or

$$p_nA^{-1} = (-1)^{n-1}(A^{n-1} - p_1A^{n-2} + p_2A^{n-3} - \cdots (-1)^{n-1}p_{n-1}I). \tag{6}$$

Since A is non singular and p_n equals the value of the determinant of A, equation (6) can be solved for A^{-1} in terms of the powers of A. Thus, if the numerical values for the p_r's are available the inverse matrix A^{-1} of A can be calculated from equation (6) by raising A to the appropriate powers, multiplying each A^{n-r-1} by p_r/p_n ($r=0, 1, 2, \cdots n-1$; $p_0=1$) and adding algebraically the matrices thus obtained. In actual practice it has been found advisable first to calculate the adjoint $\overline{A}$ of A rather than A^{-1}. The adjoint $\overline{A}$ is by definition equal to A^{-1} times the value of the determinant of A. Since p_n equals the value of the determinant of A, $\overline{A}=p_nA^{-1}$. Hence, equation (6) becomes

$$\overline{A} = (-1)^{n-1}(A^{n-1} - p_1A^{n-2} + p_2A^{n-3} - \cdots (-1)^{n-1}p_{n-1}I). \tag{7}$$

In practice the values of p_r are as a rule not known and direct calculation of them is a laborious process. Fortunately, the process of raising the matrix to powers of itself yields the numerical values p_r. This can be seen from the following considerations. The quantities p_r are symmetric functions of the roots of equation (3). Moreover, if s_r stands for the sum of the rth powers of the n roots of equation (3) then by Newton's symmetric equations[6]

$$\begin{aligned} s_1 - p_1 &\equiv 0 \\ s_2 - p_1s_1 + 2p_2 &\equiv 0 \\ s_3 - p_1s_2 + p_2s_1 - 3p_3 &\equiv 0 \\ \cdots\cdots\cdots\

knowing A, A^2, A^3, etc. it is a simple matter to obtain s_1, s_2, s_3, etc. Using the values of the s's thus obtained together with equations (8), the values of p's are easily obtainable.

It can readily be seen that the great bulk of the work is in the determination of the adjoint matrix, $\overline{A}$. Once it is determined, the division of each of its elements by the determinant p_n yields the desired inverse matrix A^{-1}. The matrix relation $A\overline{A}=p_nI$ yields the determinant p_n as elements in the principal diagonal of the matrix p_nI, and also affords a check on the computations. So also can the matrix relation $AA^{-1}=I$ be used as a checking device. The details of the foregoing method can best be explained by working through an actual example involving four variates.

In the study of 5760 eight year old boys measured in fifteen states the intercorrelations for the variates stature, chest girth, hip girth and arm length were determined. These are given in the following matrix of correlations. Since the correlation matrix is symmetrical ($r_{ij}=r_{ji}$) only the necessary $n(n+1)/2$ terms are given in the matrix

$$A=\left\|\begin{array}{cccc|c} 1.000000 & .615429 & .674646 & .852162 & 3.142237 \\ & 1.000000 & .815032 & .608666 & 3.039127 \\ & & 1.000000 & .627702 & 3.117380 \\ & & & 1.000000 & 3.088530. \end{array}\right\| \qquad \text{Check}$$

Since n, the number of variates, is equal to four, equation (7) becomes

$$\overline{A} = -A^3 + p_1A^2 - p_2A + p_3I.$$

Given A, it remains to compute A^2, A^3, p_1, p_2 and p_3.

To insure accuracy, as each row of the succeeding matrix is evolved, a check column is added, each element of which is equal to the sum of the elements appearing and understood on the particular row to which it refers.

To square A, multiply it by itself. That is, the element which lies in the ith row and jth column of A^2 is obtained by multiplying each element in the ith row of A by the corresponding element in the jth column of A, and summing the results.[7]

$$A^2=\left\|\begin{array}{cccc|c} 2.560080 & 2.299398 & 2.385790 & 2.502391 & 9.747660 \\ & 2.413504 & 2.427322 & 2.253374 & 9.393598 \\ & & 2.513434 & 2.326394 & 9.652940 \\ & & & 2.490664 & 9.572824. \end{array}\right\| \qquad \text{Check}$$

[7] *Ibid.*, p. 63.

To obtain A^3 multiply A by A^2. That is, to obtain the element which lies in the ith row and jth column of A^3, multiply each element of the ith row of A by the corresponding element of the jth column of A^2, and sum the results. The matrices are symmetrical, therefore, rows and columns are interchangeable.

$$A^3 = \left\| \begin{array}{cccc|c} 7.717202 & 7.342561 & 7.557777 & 7.581124 & 30.198667 \\ & 7.178517 & 7.360132 & 7.205486 & 29.086698 \\ & & 7.561625 & 7.414589 & 29.894124 \\ & & & 7.454941 & 29.656143. \end{array} \right\| \quad \text{Check}$$

We now solve for p_1, p_2 and p_3. To do this we utilize equations (8). Substituting $s_1 = 4.000000$, $s_2 = 9.977682$, $s_3 = 29.912285$ in these equations; we obtain $p_1 = 4.000000$, $p_2 = 3.011159$, $p_3 = .682065$.

Knowing A, A^2, A^3, p_1, p_2, p_3; we need only multiply A^3 by -1, A^2 by p_1, A by p_2 and I by p_3. This is readily accomplished since the multiplication of a matrix by a scalar is a matrix each of whose elements is the scalar times the corresponding elements of the matrix.[8]

$$-A^3 = \left\| \begin{array}{cccc|c} -7.717202 & -7.342561 & -7.557777 & -7.581124 & -30.198667 \\ & -7.178517 & -7.360132 & -7.205486 & -29.086698 \\ & & -7.561625 & -7.414589 & -29.894124 \\ & & & -7.454941 & -29.656143, \end{array} \right\| \quad \text{Check}$$

$$p_1 A^2 = \left\| \begin{array}{cccc|c} 10.240320 & 9.197592 & 9.543160 & 10.009564 & 38.990636 \\ & 9.654016 & 9.709288 & 9.013496 & 37.574392 \\ & & 10.053736 & 9.305576 & 38.611760 \\ & & & 9.962656 & 38.291292, \end{array} \right\|$$

$$-p_2 A = \left\| \begin{array}{cccc|c} -3.011159 & -1.853155 & -2.031466 & -2.565995 & -9.461775 \\ & -3.011159 & -2.454191 & -1.832790 & -9.151295 \\ & & -3.011159 & -1.890111 & -9.386927 \\ & & & -3.011159 & -9.300055, \end{array} \right\|$$

$$p_3 I = \left\| \begin{array}{cccc|c} .682065 & & & & .682065 \\ & .682065 & & & .682065 \\ & & .682065 & & .682065 \\ & & & .682065 & .682065 \end{array} \right\|$$

[8] *Ibid.*, p. 62.

To arrive at $\overline{A}$ it remains simply to add the corresponding elements of the matrices $-A^3$, p_1A^2, $-p_2A$ and p_3I.

$$\overline{A} = \left\| \begin{array}{cccc|c} .194024 & .001876 & -.046083 & -.137555 & .012263 \\ & .146405 & -.105035 & -.024780 & .018460 \\ & & .163017 & .000876 & .012776 \\ & & & .178621 & .017165. \end{array} \right\|$$

Finally, to obtain the desired inverse matrix A^{-1}, we divide $\overline{A}$ by p_4. To obtain p_4 we make use of the matrix relation $A\overline{A} = Ip_n$. Multiplying A by $\overline{A}$ we get[9]

$$A\overline{A} = \left\| \begin{array}{cccc} .046870 & .000000 & .000001 & .000000 \\ & .046870 & .000001 & -.000001 \\ & & .046870 & .000001 \\ & & & .046870 \end{array} \right\|.$$

Dividing $\overline{A}$ by p_4 (i.e., 0.046870), we obtain the desired inverse matrix

$$A^{-1} = \left\| \begin{array}{cccc|c} & & & & \text{Check} \\ 4.139262 & .040026 & -.983209 & -2.934820 & .261617 \\ & 3.123640 & 2.240986 & -.538696 & .393983 \\ & & 3.478066 & .018690 & .272562 \\ & & & 3.810988 & .366162. \end{array} \right\|$$

As a final check on the calculations, we multiply A by A^{-1}.

$$AA^{-1} = \left\| \begin{array}{cccc} .999993 & -.000002 & .000014 & -.000007 \\ & .999999 & .000031 & -.000018 \\ & & 1.000005 & -.000014 \\ & & & .999978 \end{array} \right\|.$$

It can readily be seen that the principal diagonal elements are very close to unity, and the other terms close to zero.

[9] It is of interest to note that $p_4 = \frac{1}{4}(-s_4 + p_1 s_3 - p_2 s_2 + p_3 s_1)$ which in this case equals .046870. For practical reasons, however, it is perhaps better to obtain p_n from $A\overline{A}$, since to obtain p_n by Newton's formulae (8) it is necessary to compute the principal diagonal elements of A^n.

A PUNCH CARD DESIGNED TO CONTAIN WRITTEN DATA AND CODING

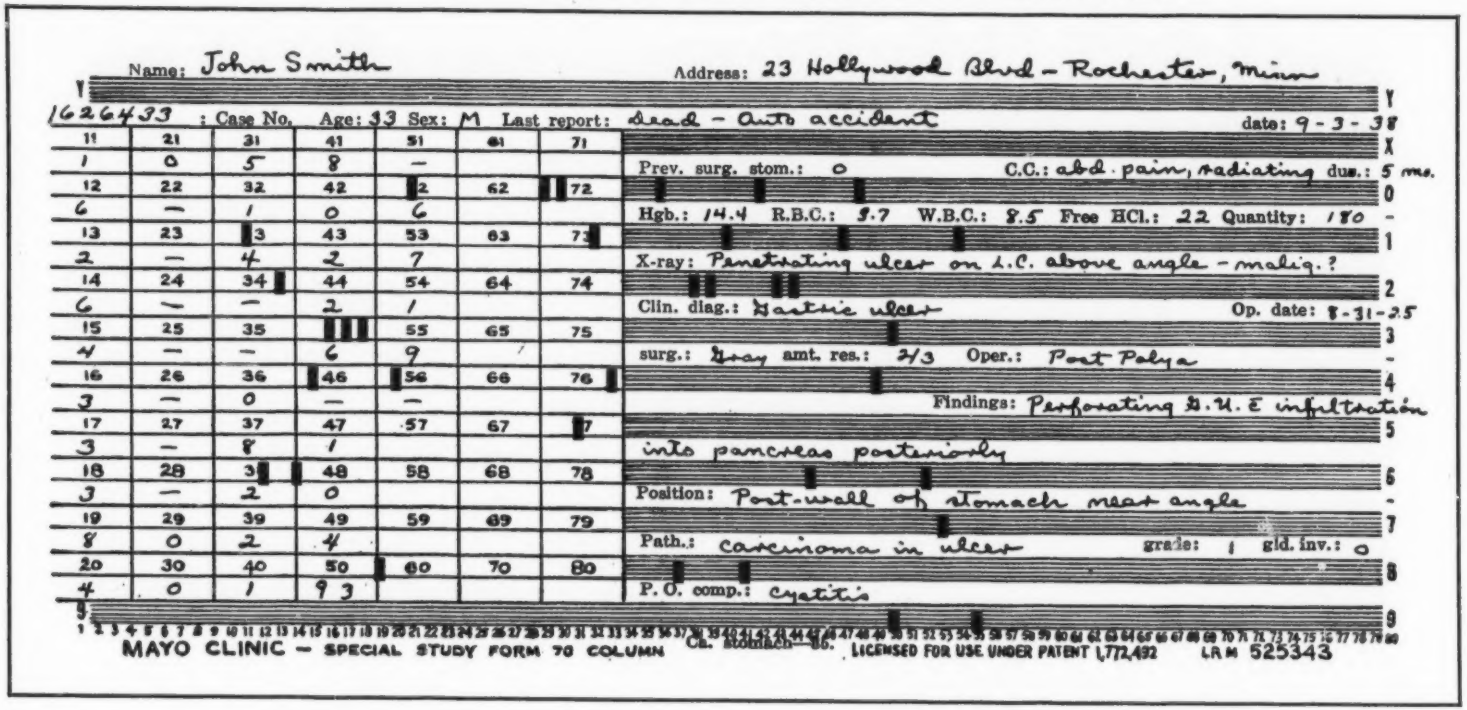
By Joseph Berkson, M.D.
Division of Biometry and Medical Statistics, Mayo Clinic

In the course of some ten years' experience with the use of Hollerith punch cards for studies requiring comparatively small numbers of cards (up to 5,000), we have developed a design of punch card that is eminently practical for certain sorts of work. The particular form now utilized is the result of evolution from a number of other forms that have been tried. It has now been stabilized for several years, so that it has had the test of a considerable amount of experience. Recent inquiries suggest that the same card may be useful to others and for this reason I am presenting it here.

The purpose of the design is to provide a card which will bear on it in writing the original data and the coding, as well as serve as a punch card. Chart I presents an example. It is seen to contain shaded areas disposed horizontally, between which are printed headings, and on the left side an area of small boxes with a column number above each box. The card is utilized as follows. There are first written against the various headings in an abbreviated form, the data obtained from whatever are the original sources of the information to be used. This part of the work is done either by a clerk or, as is the case with many of the studies conducted in this division, by the physician himself who is making a survey of case records. With the task of abstracting the data on to the cards completed, they are taken up by the Statistics Division to assist in the preparation of a code. For each item to be coded, a hand sort of a reasonably large sample is made, in order to assist in determining what groupings should be used for the categories to be coded. With the code completed, a clerk fills in the code numbers for the several columns from 11 through 80 in the spaces immediately beneath the appropriate column number in the boxes on the left side of the card. This completes the written work of the card. For punching, the card is inserted into the punch machine with a stop at the eleventh column so that the left portion of the card is exposed to the view of the punch operator. In this position the first group of column headings is exposed, so that the coding for columns 11 through 20 is in view. The clerk punches the columns in order and as this is done, the card proceeds leftward, exposing successively the coding for the remaining columns. The spacing of these boxes is such that the coding of the column being punched at any particular time is exposed to the view of the operator. After the

CHART 1

HOLLERITH PUNCH CARD DESIGNED TO CONTAIN BOTH WRITTEN DATA AND CODING


Name: John Smith
Address: 23 Hollywood Blvd - Rochester, Minn
16264 33 : Case No. Age: 33 Sex: M Last report: dead - auto accident date: 9-3-38
Prev. surg. stom.: 0 C.C.: abd. pain, radiating dus.: 5 mo.
Hgb.: 14.4 R.B.C.: 5.7 W.B.C.: 8.5 Free HCl.: 22 Quantity: 180
X-ray: Penetrating ulcer on L.C. above angle - malig.?
Clin. diag.: Gastric ulcer Op. date: 8-31-25
surg.: Gray amt. res.: 2/3 Oper.: Post Polya
Findings: Perforating D.U.E infiltration into pancreas posteriorly
Position: Post-wall of stomach near angle
Path.: carcinoma in ulcer grade: 1 gld. inv.: 0
P.O. comp.: cystitis
MAYO CLINIC — SPECIAL STUDY FORM 70 COLUMN
LICENSED FOR USE UNDER PATENT 1,772,492
LRM 525343

card has been punched, it is as readable as regards the original data and the coding as before the punching, because the punch holes will be made only in the shaded areas.

Some of the practical advantages of this sort of card are obvious, and some others which have been found in the course of actual experience are of even greater practical importance. In the use of the Hollerith punch card in the most usual form, one would have for a problem such as referred to, a sheet for the abstract of the data and perhaps another one for the codification, and the punch card would serve only for its essential purpose of sorting and tabulation. The desirability, if it is practicable, of condensing the abstract of the data and the coding on the one punch card and thus eliminating two sets of forms is obvious. The punch card utilized in this form has also an advantage for checking particular items. If, to take a simple example, the item of age is to be checked, the cards are run through the sorting machine on the appropriate columns and they are then in the order 01, 02, 03, and so forth, according to the punches. It is now an easy matter to inspect the cards to see if all those punched 01, which will be in the first group, have on the written data the same designation, and so for the successive groups. This gives a check from the original data to the punches and eliminates the check of codification. When a discrepancy is found, it is easy to locate whether the error has been made in the coding or in the punching, because the coding is also readable.

Perhaps the greatest practical advantage that we have found is one that might not appear especially impressive when described. It refers to the facilitation of making a code as well as the utilization of the cards for details so minute that it would be prohibitive to codify them. Suppose we have the task of making a code for the item "clinical diagnosis." The diagnoses as they have been abstracted on to the card may be, and in general will be, transcribed in considerable detail. We may have written "duodenal ulcer with adhesions and perforation," or "duodenal ulcer with slight hemorrhage," or "duodenal ulcer with massive hemorrhage": that is, the cards will contain a certain number of diagnoses of duodenal ulcer with a large variety of detailed modifications appended, and so for other diagnoses. At the outset of the study we cannot be certain how many of these details may be eventually useful. What we can anticipate from experience is that we shall find in the end that certain particular details for some few of the diagnoses will be extremely useful and important, but that in the main for the rest of the diagnoses there will be no particular utility in these details. With the use of an ordinary punch card in such a situation, we have the alternative of

either making a code in excessively great detail so as to be sure that what we may need later will be available in the punches, or codifying in only comparatively broad groupings with a simpler code and giving up the possibility of obtaining the details from the punch card. With the present punch card a comparatively broad grouping can be utilized in making the code. In the course of the study, if it should appear that some particular details in some particular group, let us say the cases of duodenal ulcer with hemorrhage, are required and only the main diagnosis "duodenal ulcer" has been given a code number, then it is an easy matter to pick out from among the cases of duodenal ulcers by hand sorting from the writing on the card, the cases in which there was hemorrhage. Or suppose age had been coded and punched only in decades and later data on the patients of less than fifteen years were wanted. The group "10–19 years" could be subdivided easily according to the age, which is written on the card. Thus, the card is available with a little supplementary hand sorting for the greatest detail in which the record was written.

The card as shown in the figure and as described permits the utilization of only seventy of the eighty columns (11–80). In special cases in which all eighty columns are required, the codification for the first ten columns is written in the right margin of the card and when the punching is to be done, the card is inserted for its full length in the machine, instead of only up to the eleventh column. The coding on the right margin of the card is then visible to the punch clerk, and as these columns are punched, the card will proceed leftward and the coding for the rest of the columns in the boxes will come into the view of the operator.

In actual use in the department we have available in stock a large number of the shaded cards without the headings for the items of data that are to be extracted. When a particular study is undertaken for which the card is to be utilized, the appropriate headings are inserted in the spaces between the shaded areas by either duplicating machine, mimeograph or over printing, according to the number of cards that are required. We find that it is quite easy even with mimeograph stencil to justify the inserts between the shaded parts of the card. The stock form of shaded card without the headings is provided in the regular way by the International Business Machines Company, utilizing a plate registered for the Mayo Clinic and copyrighted. The plate will be made available to others without cost if permission is first obtained. The request for permission should be directed to the Statistics Division of the Mayo Clinic.

NOTES AND DISCUSSIONS

COMMENTS ON DR. MADOW'S "NOTE ON TESTS OF DEPARTURE FROM NORMALITY" WITH SOME REMARKS CONCERNING TESTS OF SIGNIFICANCE

I

The "Note on Tests of Departure from Normality" and the promised further analysis by William G. Madow, in the September, 1940, number of this JOURNAL, pages 515–517, are welcome as a contribution toward a much needed clarification of the import of applying tests of significance. The author gives the inequalities:

$$\left|\frac{g_1}{s_{g_1}}\right| > c_\alpha \quad \text{or} \quad \left|\frac{g_2}{s_{g_2}}\right| > c_\alpha$$

and discusses, as a test of significance of departure from normality, the finding in large samples of either or both of these to be satisfied, c_α being equal to 1.96 if 0.05 is the level of significance used. He argues that the test will reject the hypothesis tested when true with a relative frequency of 0.0975 not 0.05 as required by the level of significance.

If the test is posed as Madow states it, what he says is correct. But I do not know any text that explicitly proposes the test precisely as he defines it, and I doubt that anyone applying a test to real data is likely to utilize it just that way. Judging from my own experience, a person with a real distribution to consider does not, using such tests, ask the question, "Is this distribution normal?" as an abstract general question. He will ask, "Is this distribution skew?" or "Is this distribution kurtic?" or both questions. Wishing to answer the first question, he will apply a test such as the first inequality, and, if it is fulfilled, he will be satisfied that the distribution is skew. Asking the second question, he will apply the second inequality, and, if it is fulfilled, quite independently of the first test, he will say that the distribution is kurtic; as a matter of fact he will say it is leptokurtic or platykurtic. We may assert that an individual pursuing this sort of practice will, in instances in which he is dealing with the normal curve, come falsely to the conclusion that it is skew, five times in a hundred, and under the same conditions he will come falsely to the conclusion that the distribution is kurtic, five times in a hundred. Thus the level 0.05 is adhered to.

It is, I think, of practical importance to distinguish a class of tests such as each of the inequalities discussed, from another class to which I shall presently refer. The first class, exemplified here, utilizes a function which, apart from its use in the significance test, is recognizable and employed as a "descriptive" measure. In the present examples the first inequality utilizes in the numerator a function which is a measure of "skewness," and the second inequality utilizes in the numerator a function which is a measure of "kur-

tosis." An individual will be likely to employ the first test only in so far as he considers "skewness" a relevant difference from normality in his situation, and also only in so far as he accepts as valid the particular measure of "skewness" involved; and similarly for "kurtosis." He will implicitly recognize the alternative hypotheses for which each test is relevant. Such tests are in a class with the *t*-test when this is used say, for testing the significance of the departure of an observed mean from zero. The *mean* is a descriptive measure quite apart from its utilization in a test of significance, and a test involving it is intuitively recognized as useful in differentiating an alternative, different from the hypothesis tested specifically with respect to the mean. The same is true for the *F*-test when it is used for the difference between observed standard deviations, for the standard deviation is a measure utilized for descriptive purposes, apart from its use in a test of significance.

Another class of tests is that in which functions are utilized that are not readily interpreted and not ordinarily employed as descriptive statistics. The several other tests referred to in the note of Dr. Madow tend to fall in this class, for they involve rather complex functions, which are not ordinarily utilized outside of tests of significance.[1] But the particular tests to which Madow refers are not frequently seen in application. There is, however, one test for normality falling in this class that is characteristically advanced by standard texts—the χ^2 test applied to the frequencies of a normal distribution.[2] The frequency χ^2 is not, and cannot be, a descriptive statistic of the data examined, and elsewhere I have given examples of difficulties of interpretation that arise when it is applied in a test for normality.

Tests such as each of the inequalities stated in the first paragraph suggest on their face in what situations they will be relevant: The first will naturally suggest itself as efficient and preferable to the second if the test is for "skewness," and the second will suggest itself as the more efficient and preferable where the test is for "kurtosis." In the sense that one or the other of these tests is preferable in one or the other of these situations, in what situations should one prefer the χ^2 test? From Fisher one obtains the impression that it

[1] I suggest tentatively that the two classes I have in mind can be differentiated as (1) those which in principle can be alternatively stated in terms of an estimate and its confidence interval and (2) those which cannot be so stated. For instance, if we perform a test of significance for the difference of an observed from a hypothetical mean as the *t*-test, $(\mu - \bar{x})/s_{\bar{x}} = t$, or state the estimated mean with its confidence interval as $\bar{x} \pm ts_{\bar{x}}$ and say that $\bar{x}$ is not significantly different from μ if the interval overlaps μ, we obtain the same results. The frequency χ^2 test for normality cannot be stated alternatively in terms of the confidence interval of an estimated quantity.

[2] F. E. Croxton and D. J. Cowden, *Applied General Statistics*, New York, Prentice-Hall, Inc., 1939, p. 287.

Raymond Pearl, *Introduction to Medical Biometry and Statistics*, Philadelphia, W. B. Saunders Company, 1940, p. 323.

P. R. Rider, *An Introduction to Modern Statistical Methods*, New York, John Wiley & Sons, Inc., 1939, p. 109.

L. H. C. Tippett, *The Methods of Statistics*, London, Williams & Norgate Ltd., 1931, p. 66.

A. E. Treloar, *Elements of Statistical Reasoning*, New York, John Wiley & Sons. Inc., 1939, pp. 223–225.

See also Joseph Berkson, "Some Difficulties of Interpretation Encountered in the Application of the Chi-Square Test," this JOURNAL, Vol. 33, September 1938, pp. 526–536.

should be used wherever the numbers in the sample are large enough to permit it. I believe, myself, that these questions can be clarified best in terms of the Neyman-Pearson power analysis, in which it is recognized that a test is meaningful and can be discussed intelligibly only in terms of an alternative to the hypothesis tested. I hope that when Dr. Madow presents his power analysis of tests for normality, he will include an ample consideration of the frequency χ^2 test of normality.

JOSEPH BERKSON, M.D.

Division of Biometry and Medical Statistics
Mayo Clinic

II

It seems to me that Madow's paper needs a good deal of qualification. The author states that a "frequently used technique" for testing departures from normality is to compute g_1/s_{g_1} and g_2/s_{g_2} where g_1 and g_2 are respectively $k_3/k_2^{3/2}$ and k_4/k_2^2, the k's being the well-known functions introduced into statistical analysis by R. A. Fisher, and s_{g_1} and s_{g_2} are standard deviations (exact formulae of which for samples of N have been given by Fisher); if $|g_1/s_{g_1}|$ and $|g_2/s_{g_2}|$ are either or both greater than 1.96 (the normal value for an .05 level of significance) then "it is customary to say that the hypothesis that the sample is a random sample from a normal distribution is rejected, the probability of error in that statement being α (equal to .05) if the sample is in fact a sufficiently large sample from a normal distribution."

It is, of course, true that for "sufficiently large" samples g_1/s_{g_1} and g_2/s_{g_2} are normally distributed. The exact sampling distributions of these functions are not known, but since, as R. A. Fisher has shown, the first six semi-invariants of g_1/s_{g_1} are close to their normal values zero for medium-size samples of N, the assumption of normality as a working rule for this ratio is justified. It is quite otherwise with g_2/s_{g_2}: even when N is as large as 101 the $\sqrt{B_1}$ and B_2-3 for $b_2-3(=m_4/m_2^2-3$ and very similar to $g_2)$ are 1.27 and 3.75 respectively which are far removed from the normal values of zero. Using Fisher's formulae for the moments, E. S. Pearson[1] has determined by means of a Pearson-type curve the 1 per cent and 5 per cent probability points for b_2 for large samples (from $N=100$ by intervals to $N=5{,}000$). For $N=100$ the 1 per cent probability points[2] are 2.18 and 4.39 which are significantly different from the values 1.88 and 4.00 which would be found if b_2 were assumed to be normally distributed. If b_2 is to be used as a test of normality, E. S. Pearson's tables for large samples should be used unless and until the exact frequency distribution of this function for normal random samples has been determined.

So far I am not seriously at issue with the author. I am, however, frankly incredulous that any statistician confronted with the proposition that the

[1] E. S. Pearson, *Biometrika*, XXII, 1930, pp. 239–49.

[2] I.e. the points which cut off 1 per cent of the frequency from each end.

probability of each of two events (assumed to be independent) is α could conclude that the probability of at least one was also α. One is not surprised to read in footnote (1) that "Fisher does not explicitly state that he favors following the erroneous procedure embodied in equations (1)." In footnote (2) the author says "it is noteworthy, however, that this (the statement about probability given some lines above) is rarely stated." I suggest that it is not stated because no statistician could possibly intend it. If either or both the functions g_1 and g_2 are below the level of significance for a very small probability statisticians are wont to conclude that the universe is probably not normal, *but without stating the exact probability.*

The alternative techniques which the author discusses are open to the objections that (i) they are unnecessary and (ii) they apply only when the samples are so very large that g_1 and g_2 are approximately both normally distributed and independent.

Of these objections (i) is the more important. Why should it be necessary to combine the two tests? An insignificant value of g_1 and a significant value of g_2 (working to a predetermined level of significance on a probability scale) indicates that the sample was probably drawn from a symmetrical non-normal universe. If g_1 is significant and g_2 insignificant one deduces that the universe was probably skew and, therefore, non-normal. In other words, one can deduce more about the universe by considering the test functions separately than by combining them. Of course, the author says that "there is no reason given why either of them (the alternative techniques) are good methods to use." But, if not, why give them?

In his concluding paragraph the author states that his comment would apply "to the $a, \sqrt{b_1}$ test of normality of E. S. Pearson and R. C. Geary." In any other context one would have assumed that "test" was simply a printer's error for the plural "tests." Here it assumes an almost sinister import. It is necessary to point out that, in the publication referred to,[3] tables of the 1 per cent and 5 per cent probability points were given for the function a (which represents the ratio of the mean deviation to the standard deviation calculated from the sample, the value of which tends towards $\sqrt{2/\pi}$ for normal random samples as N tends towards infinity) and for $\sqrt{b_1}$ and b_2 for normal samples, and nowhere was there the faintest suggestion of a combination of tests. The exact sampling distribution of a is not known but there is scarcely any doubt that for all values of N its sampling distribution is close to the "first approximation to the law of error."[4] On this basis a table of probability points has been computed from $N = 11$ and upwards.[5]

The function $\sqrt{b_1}$ will have a small value for all samples from symmetrical universes. If, as seems likely, symmetrically non-normal frequency distributions are of more frequent occurrence than non-normal distributions with

[3] R. C. Geary and E. S. Pearson, *Tests of Normality*, Issued by the Biometrika Office, 1938, University College, London.

[4] R. C. Geary, *Biometrika*, XXVII, 1935, pp. 310–32; XXXIII, 1936, pp. 295–305.

[5] See footnote 3.

b_2 tending towards 3 or a tending towards $\sqrt{2/\pi}$ as N tends towards infinity, then b_2 or a will provide more sensitive tests of normality than $\sqrt{b_1}$ which should be regarded rather as a test of symmetry. There is no reason to believe that a is a less sensitive test of kurtosis than b_2 to which it has a close inverse relationship[6] and it has the overwhelming advantage that it can safely be used for testing the normality of all sizes of samples. It is very easy to compute. I have been informed by M. Frèchet that its use as a test of normality was recommended by Bertrand, but I have not been able to obtain an exact reference. As far as our knowledge goes at present the function a appears to be the best general test of normality.

R. C. GEARY

Dublin

ROLAND POST FALKNER, 1866–1940

In January, 1889, only a year after the founding of this JOURNAL, a young instructor in statistics named Roland P. Falkner, who was then engaged in his first year of teaching at the University of Pennsylvania, joined the American Statistical Association. When Dr. Falkner died, on November 29, 1940, in his seventy-fifth year, he had been a member of the Association for more than half a century. He had a remarkably varied career, and as teacher, writer, editor, critic, administrator, and planner of research played an important role in stimulating interest in statistics and in the improvement of the quality and treatment of data in a number of fields.

Born in Bridgeport, Connecticut, April 14, 1866, he was graduated from the University of Pennsylvania at the age of nineteen and then went to Germany, where he studied economics at the Universities of Berlin, Leipsic and Halle. He was only twenty-two when in 1888, having received his Ph.D. degree from Halle, he became instructor in accounting and statistics at the Wharton School of Finance and Economy of the University of Pennsylvania. Three years later he was made Associate Professor of Statistics, the first person in this country, I believe, to be designated a professor of statistics. For twelve years he taught at Pennsylvania. During that period he led an unusually busy life. In 1891 he published his translation of Meitzen's *History, Theory and Technique of Statistics*. In the same year he served as statistician for the United States Senate Committee on Finance, and in 1892 as secretary of the International Monetary Conference. As one of the representatives of the United States he attended the International Congress of Hygiene and Demography at Budapest in 1894 and the meeting of the International Statistical Institute at Berne in 1895. From 1890 to 1895 he was associate editor and from 1896 to 1900 editor of the *Annals of the American Academy of Political and Social Science*. During that period also, besides writing on statistical and economic topics for other journals, he contributed

[6] E. S. Pearson, *Biometrika*, XXVII, 1935, pp. 333–52; R. C. Geary, *Biometrika*, XXVII, 1935, pp. 353–5.

to the *Quarterly Publications of the American Statistical Association* (as this JOURNAL was then called) a greater number of papers than any other person. He dealt with a variety of subjects, including statistics of crime, commodity prices, corporations, wages, and the census. His papers were of unusually high quality and had a directness, ease, and clarity of style not often found in statistical writing.

His best known and most important work during this early period was in the field of price statistics. Only one previous effort, and that a most unsatisfactory one, had been made to compile an index of commodity prices for the United States when in 1891 the Department of Labor, under Dr. Falkner's direction as statistician for the Senate Committee on Finance, made (to quote from Professor Wesley C. Mitchell's monograph on *The Making and Use of Index Numbers*) "a huge collection of price quotations, running back as far as 1840, and compiled an index number including more than 200 commodities for the years 1860 to 1891, and 85 commodities for 1840 to 1891." This index, which was known as the Falkner index, was widely quoted and stimulated interest in the making of price index numbers. The Department of Labor, largely as a result of Dr. Falkner's efforts, conducted an inquiry into the course of wholesale prices from 1890 to 1899, bringing the Falkner index, with changes in base period and method of weighting, down to the latter year; and in 1902 commenced the regular publication of the Department's (now the Bureau of Labor Statistics) index.

In 1900 Dr. Falkner left the University of Pennsylvania to become chief of the division of documents of the Library of Congress, where he remained until 1904, when he was appointed Commissioner of Education for Porto Rico. In this latter post, which he occupied until 1907, he became greatly interested in the improvement of school statistics, and for a number of years, not only while in Porto Rico, but also during the period 1908–1911 when he was in charge of the school inquiries of the United States Immigration Commission, he did a considerable amount of constructive work in that field.

It was only natural that one who was so much concerned with the quality and appropriateness of statistical data should actively interest himself in the work of the Census Bureau. As early as 1897 he contributed a long critical paper on "Crime and the Census" to the *Annals of the American Academy of Political and Social Science*, and the following year another long paper to the same journal on "The Development of the Census," the latter being an excellent historical study, supplemented by a critical and constructive discussion of the census statistics of population and of production. In 1898 he was appointed a member of a special committee of the American Economic Association to consider the scope and method of the approaching Twelfth Census, the other members of the committee being Richmond Mayo-Smith, Walter F. Willcox, Carroll D. Wright, and Davis R. Dewey. In 1903 he served as special agent of the Bureau of the Census on criminal statistics. And in 1911 and 1912 he was Assistant Director of the Census.

In 1909 Dr. Falkner interrupted his statistical activities to become chairman of the Commission of the United States to the Republic of Liberia,

and again in 1913 to be a member of the Joint Land Commission of the United States and Panama.

During the last twenty-five years of his life Dr. Falkner's work was chiefly of an editorial, critical, and advisory nature. From 1915 to 1923 he was editor, and from 1923 to 1926 director of research, for the Alexander Hamilton Institute. From 1926 until his death he was on the research staff of the National Industrial Conference Board, serving as editor, chief statistician, chairman of the Staff Economic Council and, more recently, as director of the Division of Information, Data and Records. Through the media of the Conference Board's publications he continued to be a prolific writer on a wide variety of subjects in the field of economics, including prices, wages, cost of living, immigration, international balance of payments, corporations, and estimates of national wealth, income, debt, and unemployment. Through voluminous correspondence with business men, universities and research organizations he disseminated a large amount of valuable data and interpretative comment. In addition, those who sought his counsel found that the breadth of his experience and the acuteness of his analytical faculty made him an excellent critic and a sound advisor.

The foregoing sketch of Dr. Falkner's activities can give little idea of the quality of the man himself. One had to know him to appreciate truly the independence of his mind, the clearness of his thinking, and his scholar's unswerving search for the truth—for the meaning behind the apparent facts. Although he was a keen and relentless critic, his criticism had that calm and impersonal quality so desirable in scholars; it was the truth with which he concerned himself, and not the human frailty of other scholars. He had a remarkable power of exposition, not only in his writing but equally in his discourse. During his long career he remained a true teacher, always interested in helping those who worked with him, by illuminating explanation, advice, and criticism, which were aimed at stimulating and encouraging independent thinking and accomplishment. His best memorial is the affection and high regard of those who were thus his beneficiaries and whose association with him remains to them a treasured possession.

F. LESLIE HAYFORD

SUMNER B. PEARMAIN, 1859–1941

Sumner Bass Pearmain, a member of the American Statistical Association for 43 years and Treasurer of the Association 1904–1919, died at his home in Boston, Massachusetts, on May 1, 1941. He was born in Boston on December 24, 1859, son of William Robert Pearmain and Cordelia Miller Smith, and he was a direct descendant of John Alden and Ann Hutchinson of colonial fame. He was graduated at Harvard College in the class of 1883.

For many years Mr. Pearmain was head of the firm of Pearmain and Brooks, stock brokers in Boston, and at one time he was a member of the Board of Governors of the Boston Stock Exchange. He also gave a series of lectures on economics at the Massachusetts Institute of Technology and at Dartmouth College.

During the later years of his life, Mr. Pearmain withdrew somewhat from his business activities, and traveled quite extensively with Mrs. Pearmain in foreign lands.

ROSWELL F. PHELPS

JOSIAH CHARLES STAMP, 1880–1941

Lord Stamp, an honorary member of the American Statistical Association, was killed in his home near London, April 17, 1941. To his many friends on both sides of the Atlantic he was better known as Sir Josiah for he was not raised to the peerage until three years before his death. For the last two years of his life he had been economic adviser to the British Government.

His connection with the American Statistical Association which was slight and indirect came mainly as a result of his membership in the International Statistical Institute to which he had been elected soon after the end of the first World War. The first meeting of the Institute at which he was present, however, was ten years later when the Royal Statistical Society celebrated its centenary in London where the Institute had originated fifty years before. At that time Sir Josiah was instrumental in persuading the Prince of Wales to accept the honorary presidency of the Royal Statistical Society and to give its guests a welcoming address which was followed on the same afternoon by a technical one from Stamp on "Estimating National Income."

American statists who attended were planning even then for a worthy centenary meeting five years later of the American Statistical Association, eldest daughter of the London Society, and it was hoped that the Institute would be a guest at the American as it had been at the English centenary. But its German president was handicapped by an imperfect knowledge of spoken English, a difficulty which would have embarrassed him in Washington even more than it did in London. Although the Institute's officers were elected every two years it had been customary to reelect every member willing to stand. But a committee to make suggestions for revising the constitution had recently proposed a process of rotation in office and although the proposal had been rejected it had led the French President to retire after a service of eight years. Plans were laid at London, therefore, to secure the early retirement of the German President also and the election of Sir Josiah in time for the projected American meeting in 1939. The first step was taken at Athens in 1936 when the Treasurer, Professor Bowley, declined to stand for reelection and suggested Sir Josiah as his successor. At that time a Belgian was chosen president but he, like his predecessor, lacked the linguistic equipment almost necessary at a session on this side of the Atlantic. The American members then hoped that Sir Josiah would become president two years later, a hope which was blasted only when the Prague session in September 1938 was cut short by the German threat to Czechoslovakia before officers could be elected.

An Englishman who knew Sir Josiah far better than I has described him thus: "Stamp was blessed at once with intellectual powers of outstanding quality, an entirely virtuous mind, to which nothing mattered but the

furtherance of the common good, amazing energy that made him ready to undertake any task that came his way, a robust physique that enabled him to tax this energy unscathed, a cheerful temperament that nothing could ruffle or fluster, and a strong leaven of practical common sense. This rare combination of gifts gave him a position of commanding eminence as an economic authority whose always disinterested judgment could be called upon by all who needed it, at a time when economic problems were the chief anxiety of a distracted world, and also secured for him a high degree of material success, the respect and affection of all who met him, and domestic happiness unsoured by the appalling problems with which he had to grapple . . . He was a splendid example of British Puritanism at its best, and his modesty and keen sense of humour saved him from any touch of that unctuousness which sometimes spoils the effectiveness of a too earnest conscience. With all his brilliant gifts, sturdy common sense was always his strong suit . . . During his career Stamp did a giant's work. All those who heard him and marked his square strong figure, his round capacious head, his aggressive moustache and bulldog expression belied by his kindly and benevolent eyes and were puzzled by his agility of thought and torrent of speech felt that they had been happy in beholding a man who was at once supremely good and immensely clever."

WALTER F. WILLCOX

GUSZTAV THIRRING, 1861–1941

The American Statistical Association has recently learned of the death of one of its Honorary Members, Dr. Gusztav Thirring, in Budapest, Hungary, on March 31, 1941.

Dr. Thirring for years was an important figure in public life in his native land. He received the degree of Doctor of Philosophy from the University of Budapest. In 1894 he became Deputy Director and from 1906 to 1926 was Director of the Budapest Municipal Statistical Office. In 1906 he was called to an associate professorship at the University of Budapest and held that chair until his retirement from active work. Under the monarchy he was an Hungarian Royal privy-councillor.

Many honors were conferred upon him, among them honorary citizenship in the free royal city of Sopron, vice-presidency (1923) and presidency (1925–31) of the Hungarian Statistical Society, vice-presidency of the Hungarian Geographical Society. In 1903 he was elected to membership in the Institut International de Statistique and in 1937 to honorary membership in that organization. He was a member of the Hungarian Academy of Science and honorary member of the German and Hungarian Statistical Societies. In December 1929 he was elected to honorary membership in the American Statistical Association.

Dr. Thirring was the author of several books. Most of these were statistical documents dealing with the city of Budapest and with the population of Hungary. He also published periodical articles in the field of population.

FRANK A. ROSS

BOOK REVIEWS

GLENN E. MCLAUGHLIN
Review Editor

The Reciprocal Trade Agreements Program, by Grace Beckett. New York: The Columbia University Press. 1941. xv, 143 pp.

This small book is a condensation of the writer's doctoral dissertation, dealing principally with one of its three sections. It begins with a brief history of the conditions preceding and the steps leading up to the adoption of the Trade Agreements Act in 1934, including the precedents for granting the Executive authority with regard to tariff matters. It then describes the procedure of framing trade agreements and discusses certain general problems, including that of "chief supplier," that of reclassification, and that of customs quotas. The next chapter describes the general provisions of the agreements, such as those mutually assuring most-favored-nation treatment, the so-called "escape clauses" (the author does not use this expression), and those regarding quantitative restriction of imports.

The remaining three chapters deal with the concessions obtained by the United States in the agreements, the concessions granted by this country, and the effects of the agreements on actual trade. Unfortunately, in these chapters, which are obviously the most important, the author has overlooked much significant published material, notably material given out by the Tariff Commission.

The importance of the concessions obtained from foreign countries is, as the author points out, more difficult to assess than that of concessions granted to them. She seems, however, to have underrated what was achieved for United States exports. For example, she speaks of "mere" bindings of free entry or existing duties, although elsewhere she does suggest, with scarcely adequate emphasis, that the agreements probably forestalled what would have been worse conditions. She says that wheat was bound duty-free by the United Kingdom, whereas in fact the former duty of about 5 cents per bushel was taken off. The many and marked reductions in duty made by Canada are not adequately described; the Tariff Commission's analysis of these, including statistics showing changes in trade under the first Canadian agreement, has apparently been overlooked.

The chapter concerning concessions granted by this country does not make use of the large body of information compiled by the Tariff Commission on this subject, which, among other analyses, shows for each tariff schedule the proportion of total imports on which concessions were made, for each schedule the average reduction in the rates of duty, and for all schedules combined the value of the trade in commodities affected by different percentages of reduction in duty. The author merely lists the number of concessions under each schedule. Moreover, in her paragraphs listing major in-

dividual concessions, the author fails to mention that on live cattle, perhaps the most important of those granted on agricultural products, or that on sugar, or those on several other major items. She mentions the binding of free entry of newsprint paper and mechanical wood pulp but not the binding of pulpwood and chemical wood pulp, both of which greatly outrank mechanical pulp in our import trade. The author's general conclusion, that the concessions granted by the United States "are probably, in most cases, as substantial as could, with prudence, have been granted" may be sound enough, but it tends to minimize the liberalization of United States tariff policy under the agreements program.

In her last chapter the author points out, very soundly, the technical difficulties of interpreting the results of trade agreements statistically. She has, however, scarcely made adequate use of the somewhat unsatisfactory statistics bearing on the subject. Thus she says that the trade of the United States since the program was adopted "has increased a little more than world trade has increased," whereas her table shows marked gains in our proportion of world trade, our share in exports rising from 11.5 per cent in 1934 to 14.0 per cent in 1938, and our share in imports rising from 8.3 per cent in 1934 to 11.3 per cent in 1937 (the decline to 8.2 per cent in 1938 has little significance in view of the sharp temporary recession of business in this country in that year). The author is right, of course, in pointing out that other causes contributed to these gains. The three tables which aim to summarize the statistics of United States trade with agreement and nonagreement countries are exceedingly confusing in form. The difficulty arises from the attempt, in a single table, to show the expansion of trade with agreement countries resulting from the making of new agreements, and to compare trade with agreement countries before and after the agreements. The author's general conclusion is that trade with agreement countries has increased more than trade with other countries, and that trade in articles on which concessions were made on either side has increased more than trade in nonconcession articles. The statistics support this conclusion fairly well. The author rightly points out that the extent to which the concessions made in the agreements have contributed to these results cannot be determined.

In the whole, Miss Beckett's book is a useful contribution, but there remains to be written a thoroughgoing description of the trade agreements and discussion of their results. Unfortunately, the war broke out too soon to give the student an adequate statistical basis for analyzing results.

E. DANA DURAND

U. S. TARIFF COMMISSION

Research Methods in Public Administration, by John M. Pfiffner. New York: The Ronald Press Company. 1940. xv, 447 pp. $4.50.

The book embodies the results of the author's systematic survey and analysis of the practices of governmental agencies, privately controlled re-

search bureaus, and individuals in investigation and research. It is a textbook for college graduates who are preparing themselves for employment as junior staff members in research agencies. It is a systematic discussion at the elementary level, and leaves little to the imagination. It makes no attempt, however, to take the discussion of any subject beyond the introductory point. The chapter on the "Use of the Quantitative Method" does not teach statistical method, but in the words of the author, the chapter's sole purpose "is to persuade students of public administration to study statistics." With that purpose the reader of this JOURNAL will surely have no quarrel.

The book is avowedly concerned with research touching the management questions in administration. For example, there is a chapter on "public personnel classification." It is also concerned largely with the management and conduct of such investigations rather than with the techniques of thought involved. It deals with the organization of work, the procedures to be employed, the personnel to participate and kindred questions.

The volume should aid young readers to visualize the activities before them in organized research, the successive steps in a research project, and the situations which will arise. Persons already active in research will be interested chiefly in the comparative resume of the methods of their fellow practitioners. The author's habit of raising questions of policy should be stimulating to both novices and initiates.

Professor Pfiffner's undertaking was ambitious and difficult, and it is clear that the work has been done with remarkable patience. Such adverse criticism as the book arouses is apt to center on its plan and purpose. Can an elementary discussion of research escape being superficial? Is the management of research peculiar enough to warrant separating it from other problems of management? Is it wise to separate procedures from the substantive problems and methods of research? In the opinion of the reviewer the product justifies the effort, the job was worth doing, and it has been done with commendable thoroughness.

GEORGE A. GRAHAM

Princeton University

Economic Factors Affecting Industrial Relations Policy in National Defense, by Sumner H. Slichter. New York: Industrial Relations Counselors Inc. Monograph No. 6. 1941. 112 pp. $1.50.

This study is a reformulation and expansion of a shorter monograph by the same author, *Economic Factors Affecting Industrial Relations Policy in the War Period*, published late in 1939. The collapse of Continental resistance to Germany, with the resulting increase in British demands for war goods and the tremendous defense program entered upon by the United States so changed the setting of the problem and the seriousness with which business must face it that the present volume is essentially a completely new study.

The first two chapters give a brief summary of the magnitude of the de-

fense program as it stood in the spring of 1941, and the next five chapters give brief estimates of the probable effects of the war and the defense program on United States exports and on the direct and indirect demand for labor. Chapter VII considers the problem of "Guns vs. Butter" and concludes that it is unlikely that the demands of the defense program and aid to Britain can be met without serious curtailment in the output of many civilian articles and the prevention of a rise in the general standard of living. Professor Slichter estimates that an increase of defense production to $20 billion a year by July 1942 (which now seems likely) would probably even require a reduction in the standard of living below the levels of July 1940.

A consideration of the question of labor shortages leads to the conclusion that shortages of some types of labor skills are already apparent, and that an increase in the demands on industry and a speeding-up of delivery dates would lead to wide-spread shortages. Because of this situation, a rapid expansion of apprenticeship and upgrading and a change in hiring and retirement policies with respect to older workers is essential to industry.

Chapters XIII, XIV, and XV consider the problems of wage and hours policy and the effects of the Public Contracts Act. The necessity for amending the hours provisions of the Fair Labor Standards Act, the study concludes, depends on the effectiveness of price controls and the willingness of the community to accept a reduction in the standard of living. Wage-rate increases should be limited and made on a temporary basis, such as the Canadian war-bonus plan, in so far as is possible, in order to prevent a rising spiral of prices and to avoid a distortion of price-cost relationships in the post-defense period. The Public Contracts Act comes in for criticism on the ground that it may prevent the low-paid workers from participating in government work. The problems of expanding unionization, strikes, union-employer contract provisions, and union-employer cooperation are next considered. The final chapter gives brief consideration to the shift from a defense to non-defense economy and the possible dominance of governmental controls in the readjustment period.

The whole study is compact and closely reasoned and, for such a short monograph, contains a wealth of information and constructive analysis. It should be of interest to all laymen and professionals interested in the broad field of labor-industry relations in the present situation.

Francis M. Boddy

University of Minnesota

Social and Economic Aspects of Swedish Population Movements, 1750–1933, by Dorothy Swaine Thomas. New York: The Macmillan Co. 1941. xxiii, 487 pp. $6.00.

This is a study of population movements as revealed by the thorough and detailed Swedish census system. The general significance of the work is heightened by the fact that during the period covered Sweden went through

important transformations in the agricultural system and the development from an agricultural to an industrial country.

A general and basic analysis of population changes from 1750 to 1933 is first made. This shows not only the growth of population, but also changes in legitimate and illegitimate birth rates, death rates, marriage rates, age and sex ratios and urbanization. The effect of industrialization upon these rates is then considered. The relation of emigration to business cycles in Sweden and in the United States is carefully analyzed.

This general analysis is followed by a detailed study of population changes from 1895 to 1933, by types of communities divided into six classes:—agricultural, rural-mixed (agricultural and industrial), rural-industrial, towns, rural communities shifting from one of the preceding classifications to a more industrialized one between 1905 and 1914, and rural communities making a similar shift between 1915 and 1924. These types of communities are contrasted as to rates of growth; the part played by natural increase, internal migration, and external migration in this growth; mortality; age; sex; and marital structure. Changes in these rates are analyzed in relation to Swedish and American business cycles. A similar analysis is made of internal mobility. In certain cases, the analysis is carried to further detail by contrasting northern and southern, forest and non-forest communities, and by treating Stockholm separately from other towns. Statistical techniques used are carefully explained in the appendix.

The analysis is based upon a vast amount of statistical work, painstakingly and thoroughly done. The basic figures are given in complete tables, making possible further studies by those interested. The results are graphically presented in numerous and excellent charts. Extreme care is taken to avoid the pitfalls of statistical analysis, and to guard against drawing unwarranted conclusions from correlations. Tentative conclusions are subject to most careful and critical examination. Correlations of general sociological significance are revealed between the demographic factors, and urbanization, industrialization, depression, and prosperity.

More attention might have been given to showing the interrelation and significance of some of the phenomena discussed. A more complete summary, assembling the large number of details developed in regard to changes in the Swedish population into a comprehensive picture, would be illuminating.

W. LEWIS ABBOTT

Colorado College

Urban Planning and Land Policies, by National Resources Committee, Report of the Urbanism Committee. Volume II. Washington. 1939. v, 366 pp. $1.00.

The growing interest in planning for the rebuilding of blighted areas in our cities lends unusual importance to this monograph, which contains the best collection of illustrations of planned cities and neighborhoods that has yet

appeared. Here are presented photographs and site plans of such model communities as Radburn, N. J., and Chatham Village, Pittsburgh. A bibliography furnishes references to 100 specific cities or neighborhoods. Data from real property in inventories showing the physical condition of housing, and census figures of median rentals and single-family homes afford a basis for comparing the housing structure of different cities. The discussions of the evils of premature subdivisions and of land speculation present somewhat familiar but nevertheless sound precepts. Methods of condemnation of land for public use and the need for more extensive municipal ownership of land are carefully treated. This monograph has performed the service of bringing together in one volume widely scattered sources of data and hence it is an invaluable reference manual for the city planner or the student of urban growth and structure.

Homer Hoyt

Chicago Plan Commission

Case Studies of Consumers' Cooperatives, by H. Haines Turner. New York: Columbia University Press. 1941. 330 pp. $2.50.

Dr. Turner's study deals with two of the classic examples of successful consumer cooperation in America, the outstanding retail society in Maynard, Massachusetts, and the group of societies served by the Central Cooperative Wholesale of Superior, Wisconsin. Other authors have dismissed these two examples with a few pages each, or at most with a chapter, but Dr. Turner devotes half a book to each, reporting fully upon research much more thorough than that of most writers on cooperation in this country. Especially valuable is the new information which he gives regarding the geographic, economic, and social environments in which these societies developed. Maynard and the Superior region are not typical American communities. One must be careful in reasoning by analogy on the basis of cooperative success there.

In addition, Dr. Turner has more nearly achieved objectivity than many, perhaps most, writers in this field. We find in his book not emotionalism or sensational propagandizing but scholarly analysis.

In his final chapter, the author appraises "Cooperative Contributions and Opportunities in the United States." He covers the familiar points in a restrained and balanced fashion, concluding that the advantages usually claimed for cooperation actually are found in practice, that social considerations are as important as the strictly economic ones in determining the degree of cooperative success, and that cooperative growth is more likely to occur in the smaller cities and rural areas of this country than in urban communities. Thus, we may infer, he expects that any near-term development of consumer cooperation may be widespread geographically but probably will not embrace a sizable portion of total retail trade.

This, perhaps, is correct. However, Dr. Turner reports upon an intensive

study of a narrow field. Except in limiting this field, he makes no mention of the interesting and promising new societies growing up among essentially typical, native-born, white-collar citizens of sizable towns and cities. Here truly new ground is being broken. Important lessons may be learned from Maynard and Superior; and many of these lessons are applicable to urban, as well as to small town or rural, societies. Nevertheless, if consumer cooperatives expand markedly in the next decade or two, they may well do so among average Americans in cities or suburbs and by methods quite unorthodox according to the standards of Maynard and Superior. Dr. Turner's book makes a valuable addition to the literature of the subject and should be studied by anyone seeking a broad understanding of American cooperation; but in addition, we need a similar appraisal of the virile new coops now pioneering in the cities.

CARL N. SCHMALZ

R. H. Stearns Company, Boston

Wheat Studies of the Food Research Institute, Stanford University, California. "Wheat in the Post-Surplus Period 1900–09, with Recent Analogies and Contrasts," by Helen C. Farnsworth. Volume XVII, No. 7. April, 1941. 315–386 pp. $1.00.

The annual reviews of the world wheat market published by the Food Research Institute since 1923 are an achievement which in comprehensiveness and especially in thoroughness has no equal in agricultural literature. It was a very good idea on the part of the Institute to give a detailed year-by-year analysis of the wheat market in the years prior to those covered by the annual reviews. Since great thoroughness was desired, the idea could be realized only step by step. Dr. K. M. Bennett reviewed the years 1914–18 (*Wheat Studies*, November 1938). Dr. Farnsworth needed two issues of the same magazine (June–July 1934) to cover most of the 1890's. In Part II of the issue at hand, the same author takes up the succeeding 9 years (1898–99 to 1908–09), promising to complete this important job in another issue by analyzing the 5-year period preceding the first world war.

The 7-year interval between the two Farnsworth issues is clearly explained by the fact that "it was necessary to collect, study, and interpret a great mass of statistical data and market information for each of the crop years covered." The author states furthermore that "long after interest in the current wheat-surplus problems has subsided, these annual reviews should prove useful in analysis of other problems that may arise in connection with wheat supplies, trade, consumption, carry-over, and prices." The reviewer can only second these claims.

Part I of the study under review is mainly devoted to a comparison of the wheat-surplus period of the 1930's with that of the 1890's. One cannot suppress the feeling that, in this comparison, factors very important in a year-to-year analysis, such as stocks or timing of crops, are ascribed too great a role, at the expense of fundamental economic developments.

Dr. Farnsworth starts the comparison and indeed the whole study with the expression "history repeats itself." But does it? Is not similarity mostly on the surface while the substance is different? So far as wheat is concerned, the similarity between the world-wheat situations in the two periods compared is mainly in the fact that supplies exceeded demand and that this naturally brought about low prices and large stocks. Almost all other factors are different: Rate of population growth, then rapid, now slow; per capita consumption of wheat, then rising, now falling; import demand, then growing, now tremendous decline almost overnight; pressure of surpluses, then evenly distributed over many countries, now weighing on a few countries and therefore much more severe; way out, then acceleration of the increase in demand and perhaps also slowing down of the increase in production in a large part of the world, now a probable drastic contraction of production in a few surplus countries. Even the timing of big and small crops, discussed in detail by the author, proved different then and now.

The material presented in the study is so exhaustive that to substantiate his dissension the reviewer would not need any other data.

N. Jasny

Bureau of Agricultural Economics
U. S. Department of Agriculture

Common Stocks and Bonds as Long-Term Investments, by Leo Spurrier. Studies in Business Administration, Vol. XI, No. 2. Chicago: The University of Chicago Press. 1941. xiii, 91 pp. $1.00.

In an investigation of the relative desirability of common stocks and bonds as long-term investments, Professor Spurrier gives an interesting summary of the conclusions of about 20 studies published by other investigators in which investment results in various classes of securities are reported. He also gives the results of 16 tests of his own which include only securities of corporations having both common stocks and bonds quoted on the New York Stock Exchange. In each case the bonds thus selected compose one index and the corresponding stocks another. Periods covered are 1913–1935 and 1923–1935. Indexes are presented separately for railroads, industrials, utilities, and also in the railroads for various subgroups. The method is to assume one dollar to have been invested in each stock and each bond at the beginning of the period examined, and to compute the results of this investment over the period. The author's conclusion is that in general the evidence favors common stocks as against bonds except for periods of falling commodity prices, such as 1923–1935.

The report is somewhat lengthy, occupying 91 pages, and might have been considerably condensed without loss of essential material. In view of this, it is regrettable that lists of the securities composing Professor Spurrier's various indexes are not given, together with a tabulation of the substitutions employed where securities had to be dropped for one reason or another. This would have facilitated judgment as to the representative nature of the sam-

ples employed. For example, the adequacy of the 24 industrial common stocks selected as representative of their class for the period 1913–1935 is subject to suspicion. Even at the beginning of this period almost 100 industrial common stocks were being actively traded on the New York Stock Exchange, so that the author's sample might have been four times as large had he not limited himself to corporations having both stocks and bonds traded on the Exchange. The imposition of this restriction may also have introduced a bias since the majority of industrial corporations do not have bond issues and the common stocks of those which do are rendered more mercurial by the introduction of the leverage factor.

Regardless of the inadequacy of the author's own investigations, which he readily admits, the paper represents the best available summary of evidence as to the relative investment merit of common stocks and bonds.

ALFRED COWLES

Cowles Commission for Research in Economics

Principles of Investment, by George Herberton Evans, Jr., and George E. Barnett. Boston: Houghton Mifflin Company. 1940. vii, 225 pp. $2.25.

This is an excellent elementary introduction to the first principles—or rather, first points of view—to be considered by the investor, private or institutional. The comprehensive character of the approach within the limited space automatically forbade attempting much more than elementary treatment, and the authors have done a remarkable job in presenting a wide range of aspects in the scope of a small volume. They start out with a discussion of what yields mean and how they are arrived at, and continue by telling what should be primary knowledge about the pure interest rate and its role, about the different kinds of risks, about liquidity, "taxability" and other fundamental aspects of every investment, which are called investment qualities. Two concluding chapters deal in a short but pertinent fashion with the desirability of putting one's funds into "investment associations" (savings banks, life insurance companies, building associations and investment trusts) and with some basic issues involved in the analysis of different types of securities.

Of course, it is unavoidable that in a presentation of general rules to guide the investor commonplaces and self-evident points of view should be emphasized. But the present book distinguishes itself most advantageously from the literature of its field by the attempt to formulate the underlying principles in a logical and systematic fashion. It is the systematic feature which gives the book its character and place especially for academic purposes.

Far more difficult is to appraise its value for "practical" purposes. Much of the analysis offered by the authors involves no new contribution from the business man's angle. In part, the refinement is carried beyond the point of usefulness. Moreover, the authors virtually omit to deal with what is the

most important practical approach: the methods to gauge *in concreto* the factors which affect investments. We would breed an anemic type of financial advisors and managers if they knew only the mathematics of investment and the typology of risks, and were not prepared to discuss in analytical terms the problems affecting those risks. What is the use of knowing how to figure out the pure interest rate if one has no way of knowing how to analyze the forces (beyond vague generalities) which affect the future rate? Even more complicated problems arise in connection with the analysis of non-governmental securities.

It may be unfair to ask a book for a performance which it did not intend to give. It is characteristic, however, that the chapter devoted to security analysis deals only with such formal aspects as ratios, accounting and statistical methods, etc. Typical, too, in that connection is the authors' reliance, along conventional lines, on past earnings and other "historical" figures without offering any appraisal of them. "A basic idea of this book is that an understanding of the principles of combination is of greater value to the investor than a mastery of the techniques for making specific investment selections." This can be accepted as axiomatic only with two limitations. First, it holds only in a very basic way; a quantitative "evaluation of investment qualities" as implied by the authors is entirely impracticable. The "principle of combination" amounts therefore to little more than the generalization that all advantages and risks of an investment should be considered. Secondly, the practical application of the theory of combination, i.e., the comparison of the diverse investment aspects, pre-supposes factual knowledge (forecast) about the latter—which is the very thing the authors relegate to second place, and which can not be discussed in terms of short-cut "principles" or of mathematical appearances.

Melchior Palyi

University of Wisconsin

The Security Affiliates of National Banks, by William Nelson Peach. Baltimore: The Johns Hopkins University Studies in Historical and Political Science. 186 pp. $1.50.

The great growth of security affiliates both in number and importance came about in the period from 1916 to 1930. They were created primarily so that national banks could "take advantage of opportunities to make profitable investments and to engage in other transactions which were presented . . . in the ordinary course of their commercial banking business" in which they could not engage directly.

Though the Pujo Committee had recommended as early as 1913 "that national banks should be prohibited from directly or indirectly engaging in the underwriting and sale of corporate securities," and though "affiliates, for the most part, carried on types of business which were either expressly prohibited by statute or which the courts had declared to be *ultra vires*," Con-

gress made no attempt to regulate such activities. In 1920 the Comptroller of the Currency stated that "some affiliates had become instruments of speculation . . . and that many of the securities floated by affiliates had proved disastrous to their subscribers and . . . had damaged the credit and reputation of national banks with which they were allied." Later the office of the Comptroller took the opposite view, and from 1923 to 1927 "was the driving force behind the legislation [McFadden Act] which conferred on national banks the power to engage in a modified security business." At that time little demand arose for the regulation or control of security affiliates. Indeed, when the Glass Bill providing for the abolishment of security affiliates was introduced in the Senate on January 21, 1932, "it met with immediate opposition from the White House, the Federal Reserve Board, the Treasury Department, the Comptroller of the Currency, and the banking community." The revelations regarding many of the practices of the affiliate system which resulted from the investigation of the Senate Committee on Stock Exchange Practices so changed popular opinion, however, that the Banking Act of 1933 was readily passed, forcing the separation of commercial banks from their security affiliates.

While the author thinks that "proper regulation and control could have prevented many of the abuses connected with the affiliate system during the twenties," he adds that lack of adequate information makes it impossible to determine whether it was wiser to abolish the practice or to attempt to regulate it. He concludes: "By abolishing affiliates in 1933 Congress removed at least one source of potential danger to the quality of bank assets, and no more practicable solution of the problem can be safely proposed."

Although the study presents little new information, it does bring together for the first time in concise form much material, hitherto not readily available dealing with the growth and operation, legal aspects, and abuses of security affiliates. The study derives its chief value, however, from the fact that it is one of the first objective treatments of a subject too often handled sensationally.

KENNETH LEWIS TREFFTZS

University of Southern California

Bank Audits and Examinations (Revised Edition), by John I. Millet. New York: The Ronald Press. 1941. xx, 577 pp. $6.00.

First published in 1927, this book has been revised according to the author's preface "to provide additional data on many points," especially on the subjects of internal control and trust departments.

Internal control is a subject which can scarcely be over-emphasized for, as the author states "the only real audit or examination is one which is made continuously as the transactions take place." The addition of new chapters dealing with methods of internal control of each major asset and liability should be of great value to anyone interested either in improving present

methods or in devising new systems of internal control. The four new chapters on trust departments replacing the one in the previous edition discuss the inspection of assets, management, corporate trust, and internal control.

The revisions, however, are more than mere additions. The changed banking conditions and the shifting of emphasis during recent years have made it possible to omit parts of the earlier work, and some chapters have been condensed and rewritten as parts of other chapters. Thus we find the discussion of statement and ratio analysis, formerly treated separately, now adequately presented in the section dealing with the examiner's report.

The fundamental outline of the first edition has not been changed. Examination and auditing techniques are discussed in detail from the procedure used in preparing for an examination to the form and presentation of the final report. Specific attention is given to the methods to be followed in inspecting each asset and in verifying each liability. Repeated recommendation is made that verification come from outside sources whenever possible, the belief being expressed that "the time will come when it will be the established practice to ask each depositor to verify his account direct to the examiner." The auditing procedure for the analysis not only of assets and liabilities, but also of income and expenses is presented. Particular attention is given to the examination of the various departments often operated in conjunction with the banking business: safe deposit, title insurance and surety, foreign exchange, and trust departments.

Although some repetition occurs as it must in a detailed discussion of examination and auditing procedures for all the assets and liabilities of a bank, the book is well done by a highly qualified author, and should be valuable not only to bank examiners and auditors *per se* but also to all bank officers and directors. A study of the book would be profitable for those in small as well as large banks and undoubtedly would result in better bank management and in a material reduction of the monthly average of 50 bank embezzlements at present reported.

Kenneth Lewis Trefftzs

University of Southern California

Measurement of the Social Performance of Business, by Theodore J. Kreps. Temporary National Economic Committee. Monograph No. 7. Washington. 1940. ix, 207 pp.

This study is devoted to a measurement of the social performances of business in a series of industries, including automobiles, boots and shoes, meat packing, cotton goods, worsted goods, chemicals, electric light and power, iron and steel. The study also deals with the groups of industries from which selections were made. The groups include manufacturing, agriculture, mining, transportation and public utilities, and construction.

Social performance is discussed by answering the following questions: (1) How many workers has the industry, or group of industries, employed

in the period from 1919 to 1938? (2) What has been the trend and fluctuations of the volume of output? (3) How much, in money terms, has the industry or group of industries received from consumers in the form of income produced, value added, or value of products sold? (4) How much in consumer effort has the absorption of these funds involved? (If the funds received are divided by Carl Snyder's index of the general price level, the dividend represents consumer effort commanded.) (5) How have payroll trends and fluctuations moved since 1919? (6) What has been the trend and fluctuations of dividends plus interest?

In general, the rise of these indexes above the base period of 1923–25 was taken as a favorable social performance. However, a general comparative rating was made in the summary at the end of Chapter VII in which the several groups of industries were given a comparative rating. The average annual increment was taken to represent progress in social contribution in the case of the following indexes: production, employment, and payrolls.

The groups were further rated by reference to the following ratios: ratio of employment to production; production to consumer effort demanded; payrolls to consumer funds absorbed; payrolls to dividends and interest; and ratio of consumer funds absorbed to dividends and interest. The summations of the various ratings give a composite rating.

The author includes in the footnotes some criticisms by his collaborators. One of the most interesting of these is by Robert Nathan (page 43), as follows:

> The main question which I raise is the suitability of assigning responsibility or credit to specific industries, major industrial categories, or single companies for their economic performance. Not enough discussion and emphasis are given to the fact that no attempt is made here to blame or congratulate an industry for a bad or good record irrespective of external factors. For instance, I know no one would criticize the carriage industry in the early part of this century, or the bicycle industry during the period when automobile usage was expanding rapidly. Also, blacksmiths and livery stables showed a bad record over many years because of a trend which obviously they could not by themselves reverse or adjust themselves to. Of course this is not true of all industries and many which have poor records can well be "blamed."
>
> On the other hand, there are many companies or industries with good records which, almost of necessity, show favorable results. For instance, the automobile industry was bound to develop with any kind of advantage taken of technical developments. Whether Ford, Chrysler, or Kettering were predominantly responsible for this success or whether the nature of individual demand was the important factor is not necessarily relevant. The important fact to take note of is the interference of external factors in the progress of particular companies or industries.

It is not possible to go into detail here as to the results shown nor is there space for a detailed criticism of the method. There is certainly a vast amount

of important information presented for the use of students of economic trends.

Spurgeon Bell

National Resources Planning Board

Export Prices and Export Cartels (*Webb-Pomerene Associations*), by Milton Gilbert, Paul D. Dickens, et al. Temporary National Economic Committee. Monograph No. 6. Washington. 1940. xiii, 310 pp.

Industrial Concentration and Tariffs, by C. L. James, E. C. Welsh, and G. Arneson. Temporary National Economic Committee. Monograph No. 10. x, 326 pp.

Both of these studies deal with aspects of the relationships between our domestic economy and our international economic policies. The first study includes three reports, two prepared by the Department of Commerce, and one by the Federal Trade Commission. Part I is an analysis of the differences between domestic and export pricing policies of United States corporations. Part II is an inquiry into the extent of foreign investments in American industry. Part III is a report by the Federal Trade Commission on the operation of associations formed under the Webb-Pomerene Law.

The effort of Dr. Gilbert to resolve the problem of price discrimination between domestic and export markets commands respect, not only by virtue of the careful delineation of the difficulties inherent in the question, but also because of the clarity with which his procedure is defined. Of the results of this reconnaissance, which covers 76 "cases," it may be said, as the author recognizes, that they are indicative, but not conclusive.

The reports on direct foreign investments in American industry, and on the operation of Webb-Pomerene export associations offer valuable reference data. An exception must be taken to the discussion which accompanies the Federal Trade Commission report, as the apologia offered serves only to emphasize the need of a critique of the issues.

Turning now to what is, in a sense, the reverse of the shield, the problem of industrial concentration and tariffs is examined in a balanced, well-planned survey of 1807 products reported in the 1937 Census of Manufactures. A special sample of 317 products comprising about 90 per cent of the items in the high-concentration class of the larger group, is used to isolate as far as possible whatever nexus may exist between restrictions and the concentration of industries thus sheltered. The general conclusion is a qualified negative to the effect that there is no "close connection" between tariffs and a "high degree" of concentration. Tariffs are easily created, but not readily destroyed, and while a tariff may be in any given case the *sine qua non* of concentration, it is more generally true that they grant added immunity to industries in which other factors produce concentration. Somewhat ironically, Dr. Kreps reminds us in his foreword that the policy of tariff walls

which contribute to industrial imbalance is historically in accordance with what Henry Clay called the "American plan."

CHARLES A. WELSH

Adelphi College

Profits, Productive Activities, and New Investment, by Martin Taitel. Temporary National Economic Committee. Monograph No. 12. Washington. 1941. xx, 188 pp.

This study deals principally with three topics. First, annual measures of profits covering the period 1909–37 are presented for American corporate enterprise as a whole. These measures, derived chiefly from *Statistics of Income*, are given in dollar amounts, as percentages of net worth, and as percentages of national income originating in corporate enterprise. The tabulations cover more years than do previously available calculations based on *Statistics of Income*, are well arranged and are supplemented by a discussion of the qualifications to which the basic data are subject. Unfortunately, the author says nothing concerning such matters as the accepted methods of computing depletion, the distribution of profits in the form of officers' compensation, and certain deductions allowed life insurance companies, which must affect the interpretation of the data.

The disposal of corporate profits is the second subject covered. The author describes the division of profits between dividends and corporate savings; and attempts also to measure the amount of personal savings out of dividends and the fraction of these savings that finds its way back to corporations through stock issues, despite the questionable assumption necessarily involved in these measurements. The author is led to the conclusion that usually "the savings created but not absorbed by corporate equity accounts have acted to increase the volume of savings which must be absorbed in other areas or in other forms in order to prevent declines in the national income." He does not, however, make clear why this is an important conclusion or why bond issues and officers' salaries and bonuses, for example, are ignored in this connection.

Finally, the author turns to an analysis of the relation between rates of return on capital in individual companies and the "rates of absorption of savings" by them, the latter being measured by the percentage change in total assets, excluding changes in cash and those due to revaluation. There is a supplementary analysis, along similar lines, of the relation between the rate of return and the "rate of investment in property," i.e., the percentage change in the net property account, adjusted for revaluations. Here the study is restricted to the published corporate reports of one and two industrial divisions respectively, though mention is made of unfinished analyses of other industries. The coefficients of correlation between the factors noted are fairly low and more frequently positive than negative. The slopes of the regression lines are usually less than unity (rate of profit taken to be the in-

dependent variable), and the positions of the lines vary with changes in business conditions. The author concludes that the level of output in relation to capacity and the development of new technologies are more important determinants of the rate of investment than is the rate of profits. The small role ascribed to the profit rate is not surprising since the analysis deals with the *current* rate of profits rather than with the expected rate.

Although Mr. Taitel's conclusions and inferences are not always drawn from the data analyzed in the present report, the monograph is well worth examination.

Solomon Fabricant

National Bureau of Economic Research

Financial Characteristics of American Manufacturing Corporations, by Charles L. Merwin, Jr. Temporary National Economic Committee. Monograph No. 15. Washington. 1940. xiv, 442 pp.

As was inevitable, some of the forty odd T.N.E.C. Monographs are more substantial, scholarly reports than others. This reviewer is happy to report that Mr. Merwin, economic analyst of the Department of Commerce's National Income Division, has contributed genuinely to what is the T.N.E.C.'s probable real claim to fame: its success in amassing a wide range of information on the American economy. In so doing, Mr. Merwin takes his place with that long line of students of the financial characteristics of American corporations: Sloan, Crum, Epstein, Nerlove, Fabricant. His monograph is a welcome addition, for it covers old and new areas, in ways both familiar and novel, chiefly during the period of the thirties.

The role of government in economic research is not a settled one, but there can be no one who would question the fruitfulness of the sort of research behind Mr. Merwin's report. Because of the confidential nature of income tax returns, classification and analysis of those returns can be done most readily by a government agency. Mr. Merwin's report bristles with new facts from a relatively unworked field. Among his materials are results of the tabulations of returns of 1300 small manufacturing concerns—results that provide mortality rates as well as records of profitability and loss. Use is also made of a retabulation and industrial classification of the corporations in the Standard Statistics Company sample of 400 corporations.

The purpose of Monograph 15 was simply to describe the financial structure of manufacturing corporations, so that no "brief" or "bias" colors the report. It is concerned with facts on profits and dividends, on working capital and fixed capital of manufacturing concerns, classified by industry and size. Finally, statistics collected by S.E.C. are used to show changes in corporate funds—the source of disposable funds and the ways in which the funds were spent. This analysis is interesting and suggestive.

Mr. Merwin stoutly defends his finding that smaller firms are less profitable than are large ones, in contradiction to the conclusion of Epstein and

other writers. He rightly points out that his is not a continuing and constant sample, as was Epstein's, but a diminishing one because of business failures, and therefore might be expected to give a truer picture. But how much of the difference arises from the different periods covered, the 1920's and the 1930's, and how much from the difficulties of truly measuring profits, one cannot tell. Mr. Merwin properly admits that some part of officers' salaries of smaller concerns must be classed with profits, and the size differences are therefore less certain.

Corporation statistics are notoriously treacherous, but Mr. Merwin apparently is aware of the difficulties. He and his staff, with the help of his immediate supervisor, Robert Nathan, formerly Chief of the National Income Division, now in the Office of Production Management, have prepared a very useful monograph.

CHARLES A. BLISS

Harvard University

Study of Legal Reserve Life Insurance Companies, by Gerhard A. Gesell and Ernest J. Howe. Temporary National Economic Committee. Monograph No. 28. Washington. 1940. vii, 466 pp.

Statement on Life Insurance. Monograph No. 28-A. 1941. v, 84 pp.

Families and Their Life Insurance, by Donald H. Davenport and Gerhard A. Gesell. Monograph No. 2. 1940. x, 168 pp.

The first of these studies, Monograph No. 28, presents in report form the significant facts developed by the Insurance Section of the Securities and Exchange Commission in hearings before the Temporary National Economic Committee. The hearings extended from February 6, 1939, to March 1, 1940, and contain 3900 pages of testimony and exhibits. Additional facts presented in the report were taken from replies to seven questionnaires sent to life insurance companies as well as from published reports and manuals dealing with insurance. Monograph No. 28-A, submitted by groups representing life insurance company management, contains two criticisms of the hearings and of the findings of fact set forth in Monograph No. 28. Monograph No. 2 is a report prepared by the Securities and Exchange Commission's Insurance Section, based on data obtained in a field survey of life insurance policy holders, conducted by the Commission in the summer of 1939, with the cooperation of the Work Projects Administration. Monograph No. 2 casts much needed light on the relation of insurance owned to family insurance needs and family income. It also supplies much of the information with respect to industrial insurance relied on by the authors of Monograph No. 28.

No investigation of concentration of economic power could well overlook legal reserve life insurance companies. The 365 such companies in the United States hold assets exceeding 28 billion dollars, and have 124 million

policies outstanding aggregating some 111 billion dollars of insurance, an amount which represents sixty per cent of the life insurance in force throughout the world. The annual receipts of these life insurance companies (consisting principally of premium payments) exceeds five billion dollars or about eight per cent as much as our national income in 1938. Current investment of such amounts, almost exclusively in bonds, Monograph No. 28 declares, is a strong inducement to industry to neglect the possibilities of equity financing which in many cases might be better suited to corporate structures.

The report levies additional criticism at agency practices and the entire institution of industrial insurance. The authors hold that the drive for new business and the emphasis on growth result in poor selection of agents, inadequate training of agents, excessive turnover of agency forces, and very low compensation. All of these conditions result in improper service to the public and improper knowledge of and regard for the interests of those who purchase insurance. The criticisms of the agency system in general, the report charges, apply with peculiar force to industrial insurance, the type which concerns low-income families particularly. Here, over-selling results in excessive lapsation which contributes to the high cost of this form of insurance.

In Monograph No. 28-A, five life insurance company presidents take specific exception to the suggestion that life insurance companies consider the possibilities of common stock investment, stating that such investment is too hazardous for insurance company purposes. Exception is taken also to some of the inferences which are drawn regarding policy terminations. It is pointed out that policy termination, in certain cases, actually fulfills one of the purposes for which the insurance was purchased; namely, provision of an emergency fund which may be drawn upon in time of need.

J. Edward Hedges

Indiana University

Large-Scale Organization in the Food Industries, by A. C. Hoffman. Temporary National Economic Committee. Monograph No. 35. Washington. 1940. xi, 174 pp.

Reports of the Federal Trade Commission on Natural Gas and Natural Gas Pipe Lines in U.S.A. Agricultural Implement and Machinery Inquiry, Motor Vehicle Industry Inquiry. Monograph No. 36. 1940. xi, 275 pp.

The Motion Picture Industry—A Pattern of Control, by Daniel Bertrand, W. Duane Evans, and E. L. Blanchard. Monograph No. 43. xii, 92 pp.

This study of food industries describes how manufacturing and distribution of meat, dairy products, flour, bakery products, fruit and vegetable canning became "big business." (Sugar refining, which is also large-scale business, is not included.) The relative importance or size of the dominant companies in each of their respective industries is clearly set forth, and an honest

effort is made to appraise their monopolistic tendencies. Adequacy of appraisal distinguishes this monograph from others in the T.N.E.C. series, which are only or chiefly compilations of factual data. The chapter on Legislative and Judicial Policy is, as the author acknowledges, "somewhat cursory." A minor point—the author is probably correct but fails to present the evidence for the allegation that "financial manipulation and overcapitalization of assets have demonstrably accompanied some of the corporate expansion in this field."

In the volume prepared by the Federal Trade Commision, much of the material consists of tables, testimony, and documents (chiefly letters) pertaining to the investigation of alleged monopolistic practices in the natural gas industry. The report presents (1) a background of the size, importance and growth of the industry with major emphasis upon the pipe line division and (2) a picture of the ownership and monopolistic practices of the four dominant groups—Columbia Gas and Electric, City Service, Electric Bond and Share, and Standard Oil of New Jersey. The Commission's exceedingly brief survey of the Agricultural Implement Industry embraces concentration of control, costs, price policies, and profits. The rigid price policy in this industry is held indefensible. The last section of the report is merely a ten page memorandum indicating the extent of compliance by auto manufacturers to the Commission's recommendation to ameliorate oppressive trade practices.

The monograph on motion pictures consists of three parts: the pattern, the issues, and observations. The pattern traces the development of the motion picture industry. Although there are 110 producers today, 8 companies produce 70 per cent of the features. The major issue is briefly that the dominant companies engage in questionable practices, such as, block booking, blind selling, enforced short subjects, zoning and setting admission prices. The Consent Decree is too recent to determine how effectively the above abuses are being remedied. The conclusions make no constructive suggestions other than the inevitable and innocuous suggestion that further study is required.

EVAN B. ALDERFER

University of Pennsylvania

The Basing Point Problem. Temporary National Economic Committee. Monograph No. 42. Washington. 1941. xiv, 148 pp.

Price Discrimination in Steel, by John M. Blair and Arthur Reeside. Monograph No. 41. xiv, 54 pp.

The Basing Point Problem contains four separate monographs offered as testimony before the Temporary National Economic Committee. Two of them were prepared by the United States Steel Corporation and two by the Federal Trade Commission. Each contribution attempts to clarify the issues involved in the old controversy as to whether price competition is eliminated in an industry that uses the basing point system of pricing. These mono-

graphs undoubtedly present the most up-to-date and informative materials relating to these problems.

"Monopoly and Competition in Steel" serves as an introduction and describes the price relationships that the Federal Trade Commission "has reason to believe exist in the steel industry." A restatement of the Commission's stand, it concludes that the steel industry is "capable of reasonably efficient operation without monopoly" and that public policy should treat it as a free enterprise and not a "natural" monopoly. It is also stated once more that the Commission considers the "open f.o.b. mill price system" essential to fair competition in steel. As further reason for enforcing such practices, it is stated that "the capitalist system of free initiative is not immortal, but capable of dying and dragging down with it the system of democratic government" and that the steel industry as a "focal center of monopolistic infection . . . may well cause the death of free capitalistic industry in the United States."

The contributions of the United States Steel Corporation consist of two statements: "Some Factors in the Pricing of Steel" and "The Basing Point Method of Quoting Delivered Prices in the Steel Industry." Contained in the former are discussions of such matters as the demand and supply of steel products, characteristics of costs, and the relation of competition to profits, capacity, and costs of distribution; in the latter the operation, history, criticisms, and alternatives of the system are examined in detail. In contrast to beliefs held by many students of industry an interesting concept, based upon operating experiences of the United States Steel Corporation, is presented to the effect that "the additional cost per unit of output remains approximately the same regardless of the rate of operations, provided labor rates, prices of raw materials, etc., remain constant." It is held that conditions requisite for "perfect" price competition can not be achieved in the steel industry. It is contended, therefore, that it is unreasonable to judge competitive practices in that industry by "imaginary standards based on abstract conditions which cannot possibly be fulfilled." In support of the Corporation's position realistic and detailed examples of the operations of the basing point system are presented. Included also are maps showing the market territories which would theoretically fall to the major steel mills producing steel sheets if each mill adopted a uniform f.o.b. mill price. In the conclusion to the second paper it is stated that "the basing point practice in the steel industry is a simple method of quoting delivered prices, which results in the competition of many geographically separated steel producers at the markets for each of the diversified products of modern steel mills. It is not a price fixing medium nor does it result in high prices. It does not stifle price competition, but rather extends the benefits of such competition to all consumers."

The fourth statement, prepared by Walter B. Wooden, assistant chief counsel, and Hugh E. White, examiner, of the Federal Trade Commission, takes the form of a reply to the United States Steel Corporation's pamphlets.

Curiously enough, this statement maintains that the corporation's defense of the basing point system relies almost completely on "abstract criteria and abstruse theory." The authors claim, on the other hand, that tangible evidence of the objectionable features of the system has been available for many years and that it sustains their stand that the system "supresses price competition and is monopolistic." They proceed to cite examples of such evidence and to take issue with many statements made in the Corporation's monographs.

Price Discrimination in Steel contains analyses on a product basis of the variations in average price per ton as related to size of shipment. The materials on which the analysis was based were gathered for other purposes and are clearly inadequate for obtaining conclusive results.

MARION W. WORTHING

National Resources Planning Board

Methods of Apportionment in Congress, by Edward V. Huntington. Senate document No. 304, 76th Congress, 3d Session. Washington: United States Government Printing Office. 1940. v, 41 pp. Mimeographed Addendum 5 pp.

The Constitution stipulates that within every period of ten years "Representatives shall be apportioned among the several States according to their respective numbers, counting the whole number of persons in each State, excluding Indians not taxed, . . . but each State shall have at least one Representative." Unfortunately the relevant populations of the respective states are not integral multiples of the quantity obtained by dividing the total relevant population by the total number of representatives, so that strict proportionality, with each representative representing the same number of persons, is impossible. Since representatives with fractional vote are not permitted, the problem is how *best* to approximate strict proportionality using only whole numbers.

Leaving legal and political questions to others, Professor Huntington discusses the mathematical features of the various methods which have been used or proposed. Is a given method free from paradoxes? In what mathematical sense does it effect the *best* apportionment? Does it tend to favor states with large or small populations as compared to other methods? These and many other questions are fully answered in the present document, and the answers are illustrated with numerical examples.

Professor Huntington points out that there are exactly five methods which are free from paradoxes. As far as processes of computation are concerned, there is no basis for choice among these five, the processes being the same except that a different table of multipliers is used in each case. He discards two of these methods on grounds of artificiality and excessive bias, and discusses the remaining three in detail. These methods and their properties in brief are:

1) *Method of Harmonic Mean:* Minimizes the *absolute* difference between the "congressional districts," i.e. the number of persons per representative,

for every pair of states; similar to a method proposed in 1832 by Professor James Dean; tends to favor the smaller states in comparison with methods below.

2) *Method of Major Fractions:* Minimizes the *absolute* difference between the "individual shares," i.e. the fraction of a representative per person, for every pair of states; proposed by Professor W. F. Willcox in 1910 and used in the last apportionment; tends to favor the larger states in comparison with methods 1) and 3).

3) *Method of Equal Proportions:* Minimizes *simultaneously* the *percentage* difference between the "congressional districts" and the *percentage* difference between the "individual shares" for every pair of states; proposed by Professor Huntington in 1921; occupies a neutral position with respect to emphasis on large or small states in comparison with the methods above.

In the mimeographed addendum these three methods are applied to the 1940 census figures and yield identical apportionments with the exception of Michigan and Arkansas, the former receiving one more and the latter one less representative according to method (2) than assigned by the others. If the present size of House (435) is maintained, then whichever of these methods is used for the current apportionment, at least nine states will lose a seat in the House, at least six states will gain a seat, and one state will gain three seats, as compared with the 1930 apportionment.

Churchill Eisenhart

University of Wisconsin

The Anatomy of American Wealth, by Robert R. Doane. Harper & Brothers. 1940. xvii, 395 pp. $2.50.

Though information on the size and distribution of wealth is most important for economic and social policy, the existing factual material is imposingly inadequate. An official census of wealth was first authorized in 1850, but a detailed breakdown of the forms of physical assets was not fully reported until 1900. In the last decennial census of wealth, public debt and taxation for 1932, however, the measurement of the total size of wealth was omitted. And according to Robert Doane in his recent book, *The Anatomy of American Wealth*, there is at the present time some question whether or not such important data will again be omitted in the forthcoming census of 1942.

Besides calling attention to this shocking lack of factual information, the author attempts to fill some of the empty "factual" boxes. From public and private sources he has rounded up data for 1922, 1930, and 1938 and has provided us with a statement covering the size and fundamental categories of our physical assets classified according to their major character. Although there will be many criticisms by specialists as to what he has included or excluded in the total and under the various categories, the book stands as a useful compilation of wealth facts from widely scattered sources.

The treatment of the distribution of physical assets is less satisfactory. Setting to one side the methods and results of other scholars in the field as

being "prejudicial ceremonial analyses," he compares the increase in the number of property owners with the increase in population. This method has two great weaknesses. The first is related to the estimate of the number of property owners which is estimated from numbers of titles to assets. For example, the number of stockholders must be approximated from the number of shares, the number of owners of savings bank deposits from savings accounts, etc. In all cases there are duplications. One individual may have several banking accounts and no allowance for this is made by Doane. He bases his estimate of stockholders on previous estimates which stress that considerable duplication exists. The number of life insurance policy owners is computed on the basis of two-thirds for all ordinary life and one-half for all industrial policies in force, to allow for duplication. But no justification of these particular allowances is made.

The second weakness in the diffusion method lies in the fact that there may be a larger number of individuals or a larger proportion of the total population holding property and still be greater concentration of wealth. An increase in the number of owners relative to the growth in the population might occur because those who formerly owned one unit of the total physical assets now own only one-half unit, the other one-half unit being taken over by new owners. It is most important to know not only whether every member in the community has at least some fractional part of the total assets, but also how the wealth which is owned is distributed among the owners. For purposes of solving the monopoly problem intelligently it is also essential that we know the distribution of active control over physical assets as well as the distribution of ownership.

It is unfortunate that sources are not given for every table. They are missing from Tables XIII, XIV, XVI, XVII, and XVIIIA. But this is minor technical criticism compared to the magnitude of Mr. Doane's task which would have challenged the capacity of a governmental agency. His courage and fortitude for undertaking the task are to be admired, even if his somewhat loose-jointed style is not. Clarity would have been gained, also, by dropping some of the Hugh Johnson type of phraseology. His chapter summarizing the work of other scholars in this field during the last century is an excellent example of opinionated criticism and summary misstatement. The best evaluation of various attempts to measure wealth and its distribution is still *Income and Wealth*, Volume III, published by the National Bureau of Economic Research in 1939. Mr. Doane makes no reference to this volume.

MAXINE YAPLE SWEEZY

Vassar College

Production and Distribution Theories, the Formative Period, by George J. Stigler. New York: The Macmillan Company. 1941. vii, 392 pp. $3.50.

The author provides a comprehensive and critical review of the contributions in the limited fields of production and distribution theory of the principal architects of modern, subjective economic theory. The immediate pe-

riod considered lies between 1870 and 1895 though the productive lives of several of the economists discussed reached down to the last two decades. Each of the first eleven chapters is devoted to the work of an individual economist, Jevons, Wicksteed, Marshall, and Edgeworth in England, the Austrians, Menger, Weiser, and Boehm-Bawerk, and Walras, Wicksell, and John Bates Clark. In the final chapter, the Euler Theorem controversy is reviewed. Although the marginal productivity theory was not fully enunciated till near the end of this "formative period," the author discusses its development rather fully in the final chapter. Dr. Stigler's discussion is premised on the assumption that production and distribution may be treated "rather fully without devoting much attention to general value theory." Likewise, he postulates a stationary competitive state which, in view of the preoccupation of the theorists he discusses except for Marshall, is quite justified.

Owing to the near intellectual isolation in which the English and Continental economists were working at this time, the complementarity exhibited in their discussions of marginal utility, interest, and wage theory is striking. The interest in alternative cost theory of Jevons, Wicksteed, and Von Weiser and in general equilibrium by Walras and Edgeworth are familiar examples.

In general Dr. Stigler's criticisms appear to this reviewer judicious and his exposition accurate and painstaking. A few points of criticism may be worth raising. Marshall's treatment of derived demand is criticized for lack of generality though in view of the restrictions imposed by static theory this seems rather pedantic. However, the author rightly calls attention to the rather unrealistic use of real costs in the Marshallian wage theory.

The critique of the Boehm-Bawerk and Wicksell production period and investment period concepts on grounds of unreality seems well taken though scant attention is paid to Boehm-Bawerk's principal contribution, namely, his exposition of the functional relationship between the duration of time over which a given amount of labor is employed and the productivity of labor. On the other hand, the comments upon Clark's concept of social capital are overly generous in view of the rather high standards held up for Marshall and the Austrian economists.

In a volume containing such profuse mathematical formulae and equations, both original and reproduced, errors in symbols seem to have been held to a minimum. On page 288, equation III, replace x with t.

O. J. McDiarmid

College of William and Mary

Business Cycles: Their Nature, Cause, and Control, by James Arthur Estey. New York: Prentice-Hall. 1941. xvii, 544 pp. $5.00 (trade), $3.75 (school).

This excellent, up-to-date textbook for undergraduate courses on business cycles by Professor Estey is a welcome contribution in a field in which useable texts are rare. The book consists of three parts: description, theory, and stabilization.

Written primarily as an undergraduate textbook the material is presented

in a clear and understandable form. Even in the part on business-cycle theory Professor Estey has done a splendid job of presenting in simple language the essential elements of the difficult theories of many representative writers. Despite the fact that any comprehensive treatment of business cycles such as this must necessarily include material which is highly controversial, Professor Estey has presented these divergent points of view in such an impartial and scientific manner that the reader is quickly impressed by the author's fairness. The book covers a wide range of the recent literature on the three fields of description, theory, and control of businesss cycles, and, in general, the discussion of the subject matter is well balanced.

Despite the general excellence of the book there are three important omissions which should be pointed out. First, even though the part on business-cycle theory is intended to be representative rather than inclusive, one is disappointed by the failure to find several familiar names. Still further, there is no discussion of the recent statistical work of Tinbergen or of the recent work with model sequences. Second, the history of specific business cycles in chronological order in the United States or other countries receives practically no attention. In the whole work there are only two charts showing the movement of business activity in the United States and even these extend back only as far as 1919. There is no discussion of actual cycles in their historical setting. For beginning students unfamiliar with the actual course of business cycles it is almost as important to give them a clear picture of the actual movement of cycles as it is to give them a composite and typical picture.

A third serious omission concerns current measurement and forecasting of cyclical movement. Although the problem is admittedly difficult, the methods open to considerable criticism, and the professional results none too promising, the current measurement and forecasting of cyclical changes in business activity constitute nonetheless one of the most important problems in courses in business cycles taught from the viewpoint of business management. Forecasting is also a basic requirement for the success of methods of stabilization which call for different types of action in the successive phases of the business cycle.

PAUL BARNETT

University of Tennessee

Business-Cycle Theory in the United States, 1860–1900, by Paul Barnett. Studies in Business Administration, Vol. XI, No. 3. Chicago: The University of Chicago Press. 1941. ix, 129 pp. $1.00.

The development of thought gives current beliefs added meaning. Realization of how an idea has been booted about in the past increases humility and adds perspective. Professor Barnett has filled in the history of business-cycle thought for the latter part of the nineteenth century, a period for which no concise study was available. The judgment of the author necessarily is a

large factor in such a study, and fortunately Barnett displays an unusual grasp of the formulations now current.

Much of the writing worth considering was found to relate to single facets of the cycle problem. Attention was largely concentrated on the downturn and downswing. With few exceptions the discussions appeared as a byproduct of studies on other economic problems, where a complete explanation of the business cycle was not called for. Most of the work was published after 1880; 11 of the 14 writers Barnett considers important wrote in this period, and some of the 11 substantially modified their analyses after 1900. The explanations became much more complex toward the end of the century.

Barnett shows that there was some conception during this period of the cycle as an oscillation, contrary to the opinion of Mitchell and Hansen. Much excellent evidence is presented, but the reviewer is not convinced that Mitchell and Hansen are completely in error. For a few of the developed theories Barnett is no doubt right, but limited emphasis should be placed on brief statements made in works devoted to other subjects because of the temptation to read more into a phrase than the writer intended. It seems unlikely that a fair proportion of economists thought in terms of business oscillation or that the general public even remotely sensed the problem.

The study discusses the business-cycle ideas of individuals fairly completely within the framework of a broad theory classification. The organization of material of this kind is always puzzling, and the method chosen appears reasonably successful. Usability of the study is reduced by absence of an index.

ELMER C. BRATT

Lehigh University

American Schools in Transition, How Our Schools Adapt Their Practices to Changing Needs, by Paul R. Mort and Francis G. Cornell. New York: Teachers College, Columbia University. 1941. xxviii, 546 pp. $4.25.

The study reported in this volume is an outgrowth of a long-time interest, and numerous investigations, in the field of public school support and control. The senior author was among the pioneers in advocating, and studying, methods of equalization of educational opportunities through both state and federal aid. He has seen the theories and underlying philosophy in this field change as a result of increased experience and insights. In recent years he has come to feel that a critical element running through the multitude of studies on optimum size of school district, methods of state and federal aid, and centralization of educational control was the factor of adaptability. That is, the widespread desire to keep education "close to the people," to resist state and federal encroachments on local educational prerogatives, and to cling to the outmoded "little red schoolhouse" are all, at heart, expressions of a desire to keep public education flexible and adapted to local and changing needs.

Since adaptability is apparently an underlying objective in both popular and professional thinking the authors have undertaken to investigate the extent to which it actually exists, and to study it as a process. Such an undertaking, properly done, involves a wide attack. The present report represents the second study directly on this problem by these two authors,[1] and it is supported by a constellation of related studies already published and under way, stimulated by the senior author. The study analyzes the cycle required for the conversion of recognized need into experimental solutions and then into widespread practice. It studies the effect of a number of factors on this conversion cycle, calls attention to the role of structure and of personnel, and points out implications for education in the large.

Extended use is made of statistical techniques, particularly some of the newer ones (analysis of variance and of covariance; chi-square tests; and other tests of significance), and the volume represents a pioneer application of these methods to a large-scale study in the field of educational administration. Thus, the facilitating or retarding effects of such factors as community culture, community type, size, wealth, tax leeway, and current expenditure, are tested for statistical significance. Details of the analysis are presented in a chapter at the end, along with another chapter on research methods.

The reader is likely to be impressed with the wealth of background, insights, and the resourcefulness of the authors. Probably the aspects of the work which represent the greatest contributions are (1) the employment of adaptability as a criterion for measuring the desirability of various features of school organization and procedure, (2) the methods of indexing the many relatively intangible factors with which the study is necessarily concerned, (3) the description of the complex aspects of educational change, and (4) the practical implications, conclusions, and suggestions which are presented.

DOUGLAS E. SCATES

Duke University

An Engineers' Manual of Statistical Methods, by Leslie E. Simon. New York: John Wiley & Sons, Inc. 1941. ix, 231 pp. $2.75.

The central question treated in this book is the application of statistical methods to industrial quality control, a field pioneered by W. A. Shewhart and other Bell System statisticians some seventeen years ago. It is avowedly a book for practical men—engineers and production supervisors—who are not specialists in statistics. Although written primarily for use in instructing ordnance officers, it is an excellent working manual for anyone confronted with quality control problems. The many possible applications in defense industries make it particularly timely.

Successful applications of statistical quality control in industry usually

[1] Paul R. Mort and Francis G. Cornell, *Adaptability of Public School Systems*. New York: Bureau of Publications, Teachers College, Columbia University. 1938. 146 pp. $2.10.

require methods which can be operated on a rule-of-thumb basis by foremen and inspectors. In an appendix, Major Simon gives an illustration of such a system installed at Picatinny Arsenal. Prior to its installation, the Arsenal management was apprehensive lest this "precise and scientific system of quality control" would substantially increase inspection costs. Actually the system was operated by less than half the original force of inspectors, the quality of the product was improved and production cost reduced.

Each such system must be tailor-made to the special conditions of the situation for which it is designed. If this job of tailoring is to be done skillfully, it requires familiarity with the manufacturing processes and the personnel involved; it also requires considerably more than a rule-of-thumb acquaintance with statistical techniques. The author does a good job of stimulating the necessary reader-interest in statistical methodology by frequent critical comments on alternative methods which interrupt his explicit directions for the use of his formulas, diagrams, and nomographic charts.

One might perhaps question whether this book ought to have been called an *engineers'* manual; this title may have both good and bad results. A possible good result would be to stimulate engineers to read the book; this is true even though some important engineering applications of statistical methods (for instance in hydrology and in telephone trunking) are not mentioned. However, it would be unfortunate if the title should discourage statisticians generally from reading this work. The author's evaluation of the uses and limitations of old and new statistical theory in a variety of situations, and his special contributions on such matters as sample size and the use of the range as a measure of dispersion make this useful reading for anyone concerned with the design of experiments or the interpretation of their results.

Eugene L. Grant

Stanford University

Introduction to Mathematical Statistics, by Carl J. West. Columbus, Ohio: College Book Company. Revised Edition. 1941. 159 pp. $2.50.

This edition is almost identical to the original edition (1918) in subject matter. Many of the illustrations and exercises have been changed and the text partially rewritten. Very short chapters have been added on the significance of differences and on small sample theory.

In subject matter this text covers curve plotting with references to interpolation, smoothing, and frequency curves; averages and measures of dispersion leading up to a study of the normal curve; correlation theory including the correlation ratio, coefficient, and index, and rank, multiple, and partial correlation; moments; significance of differences; and small sample theory.

The author states in the preface that there is no real reason why the theory of statistical methods should remain in obscurity and that the necessary mathematics is largely elementary arithmetic. The presentation is quite

good for one limited to the understanding of a student with a good background of high school algebra. An attempt, only partially successful, appears to have been made to avoid the half truths so often appearing in texts as elementary as this.

FRANKLIN E. SATTERTHWAITE

Aetna Life Insurance Company

Introduction to Algebraic Theories, by A. Adrian Albert. Chicago: University of Chicago Press, 1941. viii, 137 pp. $1.75.

The purpose of this book is to fill the gap between the standard college algebra course and the abstract algebraic theories such as are treated in the author's book, *Modern Higher Algebra*. The book is built around the theory of matrices and the algebraic concepts to which matrix theory can be applied. The need for a textbook such as this in English has become quite apparent in recent years. The mathematical statistician is being asked to develop and analyze statistical designs of increasing complexity since the introduction of the analysis of variance. To carry on such research efficiently and to present the results clearly and effectively requires full use of the most powerful and elegant algebraic methods.

The study is introduced with a chapter on polynomials which covers the fundamental definitions, properties, and operations. Chapter II has a similar function with respect to matrices, determinants, and elementary transformations. Particularly valuable are the modern condensed notations and proofs here introduced. Next questions of equivalence are investigated for many general and special types of matrices and forms when they are subjected to different types of transformations. The statistician will recognize in the chapter on linear spaces an abstract treatment of his problems of independence and dependence. The treatment of polynomials with matrix coefficients gives the practical methods by which we may break up complicated statistics such as quadratic forms into fundamental independent components. The last chapter on fundamental concepts introduces the abstract theories of groups, rings, fields, and ideals such as are treated in *Modern Higher Algebra*.

The book is well written, the theorems are presented in logical order, the proofs are rigorous but never tedious, the notation is compact and clear, and extensive well chosen exercises and numerical problems are included to drive home each point.

This book should be an excellent text to give both the pure and applied mathematician training in the theory of matrices. Advanced undergraduate and beginning graduate students should find it easy to understand. For those who may not be acquainted with Professor Albert, he is one of the leading algebraists in America, having received in 1939 the Cole Prize for the most outstanding research in algebra during the previous ten years.

FRANKLIN E. SATTERTHWAITE

Aetna Life Insurance Company

Investment Banking Under the Securities and Exchange Commission, by T. Kenneth Haven. Michigan Business Studies, Vol. IX, No. 3. 1940. 154 pp.

This review will be concerned only with the statistical part of the study, disregarding the discussion of the various aspects of investment banking and SEC regulations dispersed throughout the text and summarized in Chapter 10.

Statistically, the core of the study is an analysis of 665 issues of securities with gross proceeds of about $6,600 million registered with the SEC during the years 1934 through 1937. The issues analyzed constitute 74 per cent of the number and 85 per cent of the amount of issues falling within the scope of the study, i.e., underwritten domestic new issues of more than $100,000 offered publicly, excluding securities of new promotions and investment companies and a few other less important types. The main sources of the data are the registration statements filed with the SEC. The statistics differ somewhat in coverage from those periodically released by the SEC which cover all registered issues though now permitting a segregation of most of the categories excluded by Haven. A further important difference lies in the fact that Haven's figures are based, as far as the material permits, on actual sales rather than on registered amounts; the differences, however, are immaterial for bonds and preferred stocks.

Statistical information is most abundant on cost of flotation studied in Chapters 2 and 3. Separate figures are shown for bonds, preferred stock and common stocks, cross-classified by rating—a breakdown not given in any of the SEC publications—and by years of offering. (On the other hand, Haven does not give a breakdown by size of issue or issuer or by yield as the SEC statistics do.) Total compensation of underwriters is further broken down into payment to selling groups and compensation remaining to underwriters. Statistical data are also given on underwriters' expenses paid by the issuer and on issues offered through rights. It is very commendable that the aggregate and average figures are supplemented by frequency distributions shown in graphical form. The factual results of Haven's analysis coincide closely with the tendencies disclosed in the statistics issued by the SEC both for the period 1934–1937 and later years.[1]

Statistical material—partly novel—is present but less dominant in Chapters 4 and 6 to 8 on underwriting losses, private placement of bond issues, options to underwriters, and type of underwriters' commitments.

Material on private placements is rather scanty. The only new data concern the fees paid in connection with certain privately placed issues. It is found that in 17 of 26 issues no fees were paid and that the average in the remaining 9 issues was about ½ per cent. I have the impression that both the proportion of issues on which fees were paid and the average fee in this very small sample are lower than more complete data would show.

[1] See *Annual Report* of the SEC for the year ended June 30, 1940, pp. 249, 254–56; Release No. 572; and special reports on cost of flotation for registered securities 1938–1939 (March 1941) and on cost of flotation for small issues 1925–1929 and 1935–1938 (May 1940).

The frequency of underwriters' losses is found by observing the price behavior of the issue after offering, not an entirely reliable method. It is found that, under the definitions used, 23 per cent of the bond and preferred stock issues and 13 per cent of the common stock issues with quotations available were not successfully marketed. Some detail is given in 20 cases regarding the amount of underwriters' losses.

The attempt made in Chapter 9 to determine the rate of return on invested capital in the underwriting field suffers, like other similar efforts, from the extreme scarcity of material. The author's conclusion that the return has been moderate in recent years must therefore be regarded as his personal guess rather than as a statistically founded judgment.

RAYMOND W. GOLDSMITH

Securities and Exchange Commission

Statistics of American Listed Corporations, Work Projects Administration Study sponsored by the Securities and Exchange Commission. Part I, December 1940. xi, 347 pp. Part II, December 1941. v, 370 pp. Paper.

An immense mine of information on corporate finance is available in the registration statements and annual reports filed by some 2,000 corporations listed on national securities exchanges, under the Securities Exchange Act of 1934. These statistics have been compiled since 1936 and published as a "Survey of American Listed Corporations" in three series of processed volumes: first, the individual company reports, with industry totals; second, special surveys such as *Subsidiaries of 2052 Registrants Under the Securities Exchange Act . . .* ; and third, the *Statistics of American Listed Corporations*, or overall summaries, which do not list individual corporation data.

Part I of the *Statistics* presents balance sheet, income and security data for 1961 concerns covering the fiscal year 1937 alone. These companies represent some 60 per cent of the assets of all American corporations. Manufacturers comprise 43 per cent and public utilities 40 per cent of the assets listed, whereas railway, communication, banking, insurance and investment companies are omitted because they report to other governmental agencies.

The report is divided into three sections. First are presented the characteristics of the security issues, the location of companies, the accounting firms retained, the relationship to subsidiaries and the fiscal periods used. The second section analyzes the types of capital structure, the leverage factor, and interest and preferred dividend requirements. The third section summarizes 22 financial and operating ratios. Nearly all of the 123 tables and 20 charts of the report are in the form of frequency distributions, with cross classification of firms by type of industry and size of assets, and less frequently by volume of sales or security issues, maturity dates, coupon rates, callability, type of capital structure, rate of return, or leverage.

Part II presents for each of the five years 1935 to 1939 consolidated financial statements and frequency distributions for 1495 identical corporations

reporting for the entire period. These account for 95 per cent of the assets represented by corporations reviewed in Part I. The period covered was sufficiently "normal" in business activity to have been selected as a uniform base for government index numbers.

Probably the most valuable contribution of Part II is its presentation of detailed balance sheet, income statement and surplus reconciliation data for the five years, classified by industry group, size of assets, and type of capital structure. These tables compare identical companies, since classifications are frozen as of 1937. The balance of Part II consists of frequency distributions of the firms for each year, classified by the three principal categories indicated above, and by return on invested capital, volume of sales, current ratio, and various tangible net worth ratios. Frequencies are reported both for all 1495 concerns and for 892 manufacturers alone, including 33 sub-industry groups.

Part I thus presents a detailed financial and operating cross-section of industry for a single year, whereas Part II provides a running account of financial items alone, for a somewhat smaller number of identical firms, over a five year period. The principal medium of analysis throughout is the breakdown of consolidated financial statements by means of frequency distributions and ratios. Percentage distributions are lacking. There are few interpretations or suggestions as to how the data may be used. The arithmetic means used in Part I overstate the typical values, since most of the distributions are positively skewed. Part II contains almost no measures of central tendency or dispersion. The use of medians and quartiles, for example, would provide a more continuous picture of temporal changes than do the annual frequency distributions alone. It is hoped that these reports—particularly the consolidated financial statements of Part II—will be continued, with perhaps more explanatory material and fewer sub-classifications—except in the case of industries—to expedite future publication. The series should provide unrivaled source materials and standards of comparison for the financial analyst.

William A. Spurr

U. S. Department of Commerce

PUBLICATIONS RECEIVED*

ACTUARIAL SOCIETY OF AMERICA, AND ASSOCIATION OF LIFE INSURANCE MEDICAL DIRECTORS. Supplement to Blood Pressure Study. New York. June, 1941. 22 pp.

BECKETT, GRACE. The Reciprocal Trade Agreements Program. New York: Columbia University Press. 1941. 142 pp. $2.00.

BURNS, EVELINE M. British Unemployment Programs, 1920–1938. Washington: Committee on Social Security, Social Science Research Council. 1941. xx, 385 pp. $2.75.

BUROS, OSCAR KRISEN, Editor. The Second Yearbook of Research and Statistical Methodology Books and Reviews. Highland Park, New Jersey: The Gryphon Press. 1941. xx, 383 pp. $5.00.

CALIFORNIA, UNIVERSITY OF, BUREAU OF PUBLIC ADMINISTRATION. Determining Work Loads for Professional Staff in a Public Welfare Agency, by Herbert A. Simon, William R. Divine, E. Myles Cooper, and Milton Chernin. Berkeley. 1941. xii, 94 pp. $1.00.

CROXTON, FREDERICK E. Workbook in Applied General Statistics. (Revised Edition) New York: Prentice-Hall, Inc. 1941. vii, 181 pp. $2.10.

DEAN, JOEL. Statistical Cost Functions of a Hosiery Mill. Chicago: The School of Business, University of Chicago, University of Chicago Press. 1941. ix, 116 pp. $1.00.

DENVER, UNIVERSITY OF, BUREAU OF BUSINESS AND SOCIAL RESEARCH AND THE SCHOOL OF COMMERCE, ACCOUNTS, AND FINANCE. Housing in Denver. June, 1941. Vol. 17, No. 1. iv, 56 pp. $1.00 per year.

DOWRIE, GEORGE W. and FULLER, DOUGLAS R. Investments. New York: John Wiley & Sons, Inc. 1941. viii, 631 pp. $4.00.

ELLIS, HOWARD S. Exchange Control in Central Europe. Cambridge: Harvard University Press. 1941. xiv, 413 pp. $4.00.

FREEMAN, J. F. The Basis for a General Theory of Economic Relationships and of Economic Activity. New York: The Cornwall Press, Inc. 1941. xxi, 387 pp.

HADASSAH MEDICAL ORGANIZATION. Twenty Years of Medical Service to Palestine, 1918–1938. Jerusalem, Palestine: Hadassah Medical Organization of Palestine. May 9, 1939. xxii, 223 pp. with 151 pp. of tables.

—— Report for 1939. Jerusalem, Palestine: Hadassah Medical Organization of Palestine. 1940. 46 pp.

HAGOOD, MARGARET JARMAN. Statistics for Sociologists. New York: Reynal and Hitchcock, Inc. 1941. viii, 934 pp. $4.00.

HARRIS, SEYMOUR E. Economics of Social Security. New York: McGraw-Hill Book Company, Inc. 1941. xxvi, 455 pp. $5.00.

HART, ALBERT GAILORD and ALLEN, EDWARD D. in collaboration with the Economics Staff of Iowa State College. Paying for Defense. Philadelphia: The Blakiston Company. 1941. viii, 272 pp. $2.50.

INTER AMERICAN STATISTICAL INSTITUTE. Statistical Activities of the American Nations. Edited under the Direction of the Temporary Organizing Committee of the Inter American Statistical Institute. Washington. 1941. xxxi, 842 pp. $2.00.

KOSARY, DOMINIC G. A History of Hungary. Cleveland: The Benjamin Franklin Bibliophile Society. 1941. xxxi, 482 pp. $2.75.

* Annual reports and publications presenting statistics collected at regular intervals have been omitted from this list. Some items of minor interest to statisticians have also been omitted. The contents of periodical publications are not listed, but the attention of the reader is directed to the lists of articles in current publications which are to be found in the *Revue de l'Institut International de Statistique, Journal of the Royal Statistical Society, American Economic Review, Population Index, Transactions of the Actuarial Society of America, The Record of the American Institute of Actuaries*, and *Sankhyā, The Indian Journal of Statistics*.—Editor.

LEAVITT, JOHN ANTON. The Voting Trust, A Device for Corporate Control. New York: Columbia University Press. 1941. 216 pp. $2.50.

LITCHFIELD, EDWARD H. Voting Behavior in a Metropolitan Area. Ann Arbor: University of Michigan Press. The Bureau of Government. Michigan Governmental Studies, No. 7. 1941. vii, 93 pp.

LOFTUS, JOHN A. Investment Management. Baltimore: The Johns Hopkins University Press. The Johns Hopkins University Studies in Historical and Political Science. Series LIX, Number 1. 1941. 136 pp. $1.50.

MARLIO, LOUIS. A Short War Through American Industrial Superiority. Washington: The Brookings Institution. 1941. vii, 51 pp. 25 cents.

MILLET, JOHN I. Bank Audits and Examination. (Revised Edition) New York: The Ronald Press Company. 1941. xx, 577 pp. $6.00.

MODE, ELMER B. The Elements of Statistics. New York: Prentice-Hall, Inc. 1941. xvi, 378 pp. $3.50.

NATIONAL INDUSTRIAL CONFERENCE BOARD. The Economic Almanac for 1941–42. New York: National Industrial Conference Board. 1941. x, 464 pp. $4.00.

—— Essential Facts for Fiscal Policy. New York: National Industrial Conference Board, Inc. July, 1941. 135 pp.

NORGREN, PAUL H. The Swedish Collective Bargaining System. Cambridge, Massachusetts: Harvard University Press. 1941. xv, 339 pp. $3.50.

PEACH, W. NELSON. The Security Affiliates of National Banks. Baltimore: The Johns Hopkins University Press. Johns Hopkins University Studies in Historical and Political Science, Series LVIII, No. 3. 1941. 186 pp. $1.50.

PHILIP, MAXIMILIAN. The Principles of Financial and Statistical Mathematics. (Revised Edition) New York: Prentice-Hall, Inc. xvi, 335 pp. $3.50.

PUFFER, CLAUDE E. Air Transportation. Philadelphia: The Blakiston Company. 1941. xxiv, 675 pp. $3.75.

SALERA, VIRGIL. Exchange Control and the Argentine Market. New York: Columbia University Press. 1941. 283 pp. $3.50.

SCHULTZ, THEODORE W., assisted by WITT, LAWRENCE W. Training and Recruiting of Personnel in the Rural Social Studies. Washington: American Council on Education. 1941. xiii, 340 pp. $3.00.

SCHWARZ, SANFORD. Research in International Economics by Federal Agencies. New York: Columbia University Press. 1941. xxxix, 357 pp. $1.75.

SHEPHERD, GEOFFREY S. Agricultural Price Analysis. Ames, Iowa: The Iowa State College Press. 1941. viii, 402 pp. $3.75.

SOCIAL SCIENCE RESEARCH COUNCIL. Research in Public Budget Administration. An Outline of Suggested Research Topics. New York: Committee on Public Administration of the Social Science Research Council. 1941. 37 pp. 50 cents.

SPECIAL LIBRARIES ASSOCIATION. Directory of Microfilm Sources Including Photostat Service. Compiled by Ross C. Cibella. New York. 1941. vii, 56 pp. 75 cents.

SWEEZY, MAXINE Y. The Structure of the Nazi Economy. Cambridge: Harvard University Press. 1941. xvi, 225 pp. $3.00.

TURNER, H. HAINES. Case Studies of Consumers' Cooperatives. New York: Columbia University Press. 1941. 330 pp. $2.50.

UNDERWOOD, RALPH S. and SPARKS, FRED W. Living Mathematics. New York: McGraw-Hill Book Company, Inc. 1940. ix, 365 pp. $3.00.

VERNON, RAYMOND. The Regulation of Stock Exchange Members. New York: Columbia University Press. 1941. xvi, 152 pp. $2.00.

WENDT, PAUL FRANCIS. The Classification of Financial Experience of the Customers of a Typical New York Stock Exchange Firm from 1933 to 1938. Maryville, Tennessee. 1941. xxii, 269 pp. $2.00.

Government Publications

DEPARTMENT OF AGRICULTURE. BUREAU OF HOME ECONOMICS (in cooperation with the Work Projects Administration). Family Expenditures for Housing and Household Operation, Five Regions, by Hazel Kyrk, Day Monroe, Kathryn Cronister, and Margaret Perry. Miscellaneous Publication No. 432. 1941. v, 244 pp. 25 cents.

——— Family Expenditures for Clothing, Five Regions, by Day Monroe, Maryland Y. Pennell, Elizabeth Phelps, June C. Hopper, and Helen Hollingsworth. Miscellaneous Publication No. 428. 1941. iv, 387 pp. 30 cents.

——— Family Expenditures for Furnishings and Equipment, Five Regions, by Day Monroe, Helen Hollingsworth, Margaret Perry, and Maryland Y. Pennell. Miscellaneous Publication No. 436. 1941. v, 212 pp. 25 cents.

——— Family Food Consumption and Dietary Levels, Five Regions, by Hazel K. Stiebeling, Day Monroe, Callie M. Coons, Esther F. Phipard, and Faith Clark. Miscellaneous Publication No. 405. 1941. vi, 393 pp. 35 cents.

COMMODITY EXCHANGE ADMINISTRATION. Trading in Wool Top Futures. Circular No. 604. August, 1941. 111 pp. 15 cents.

FEDERAL TRADE COMMISSION. Industrial Corporation Reports. Bolt, Nut, Rivet, and Screw-Machine Manufacturing Corporations; Cement Manufacturing Corporations; Cranes, Dredging, Excavating and Road-Building Machinery Manufacturing Corporations; Distilled Liquors Manufacturing Corporations; Drugs and Medicines Manufacturing Corporations; Furniture Manufacturing Corporations; Lumber and Timber Products Corporations; Plumbers' Supplies Manufacturing Corporations; Summary by Industry Groups; Tanned, Curried, and Finished Leather Manufacturing Corporations; and Textile Dyeing and Finishing (Except Woolen and Worsted) Corporations.

TARIFF COMMISSION. Italian Commercial Policy and Foreign Trade, 1922–1940. Report No. 142, Second Series. 1941. xiii, 284 pp. 30 cents.

—— Latin America as a Source of Strategic and Other Essential Materials. Report No. 144, Second Series. 1941. ix, 397 pp. 50 cents.

—— Regulation of Imports by Executive Action in Countries with Independent Tariff Jurisdiction. Miscellaneous Series. 1941. vi, 106 pp. 15 cents.

TEMPORARY NATIONAL ECONOMIC COMMITTEE ON THE CONCENTRATION OF ECONOMIC POWER IN THE UNITED STATES. Review and Criticism on Behalf of Standard Oil Co. (New Jersey) and Sun Oil Co. of Monograph No. 39 with Rejoinder by Monograph Author. Monograph No. 39-A. 1941. vi, 96 pp.

—— Final Report of the Executive Secretary to the Temporary National Economic Committee on the Concentration of Economic Power in the United States. 1941. xv, 435 pp.

CALIFORNIA STATE CHAMBER OF COMMERCE, RESEARCH DEPARTMENT (with the assistance of Work Projects Administration). Handbook of Sources of Economic Data Pertaining to California. San Francisco. 1941. 232 pp. $1.00.

STATE OF NEW YORK, DEPARTMENT OF LABOR. Partial Unemployment. New York City: Bureau of Research and Statistics, Division of Placement and Unemployment Insurance. Placement and Unemployment Insurance Series Special Bulletin No. 3. 1941. xiv, 311 pp. $1.75.

—— Help-Wanted Advertising, New York City, 1939–1940. New York City: Bureau of Research and Statistics, Division of Placement and Unemployment Insurance, prepared with the assistance of Occupational Studies Project, Work Projects Administration of New York City. Miscellaneous Series. Statistics Report No. 3. 1941. iv, 103 pp.

GETTING ON WITH THE JOB

THE nation's defense today calls for machines—machines on land, on sea and in the air. Output in greater and greater quantity of planes, guns, tanks, ships and hundreds of other materials in the shortest possible time is the all-important goal.

The job calls for a yearly production of military and naval equipment alone approximating in magnitude America's greatest production of durable goods of all kinds in any past year.

In assuming this gigantic task, American industry faced many difficulties—encountered unusual problems of many kinds. For arms production lies in a field with which our peaceful nation's shops and factories had not been familiar.

But problems are being met, difficulties overcome, and defense production is forging ahead from week to week. General Motors deliveries of defense materials, for example, will have *increased nearly four-fold in the period of one year*—from $43,700,000 in the fourth quarter of 1940 to an estimated $160,000,000 for the fourth quarter of 1941.

A tremendous part of the job of building the machines and equipment for defense still remains. Along with industry as a whole, we in General Motors are devoting our best efforts to the end that this job may be well done.

GENERAL MOTORS

CHEVROLET · PONTIAC · OLDSMOBILE · BUICK · CADILLAC

Please mention the Journal of the AMERICAN STATISTICAL ASSOCIATION *in writing advertisers*

OFFICERS AND EDITORS FOR THE YEAR 1941

President

WINFIELD W. RIEFLER
The Institute for Advanced Study
Princeton, New Jersey

Directors

W. LEONARD CRUM
Harvard University
Cambridge, Massachusetts

LOWELL J. REED
The Johns Hopkins University
Baltimore, Maryland

O. C. STINE
U. S. Department of Agriculture
Washington, D. C.

F. LESLIE HAYFORD
General Motors Building
New York City

FREDERICK F. STEPHAN
Cornell University
Ithaca, New York

WILLARD L. THORP
Trustee, Associated Gas and Electric Corporation
New York City

Vice-Presidents

MEREDITH B. GIVENS
New York State Department of Labor
New York City

Collection and Classification of Data, and Administration of Statistical Agencies

ROBERT R. NATHAN
Office of Production Management
Washington, D. C.

Facts and Methods Bearing upon Economics and Economic Theory

HAROLD HOTELLING
Columbia University
New York City

Statistical and Actuarial Methods and Technique, and the Teaching of Statistics

JOHN W. SCOVILLE
Chrysler Corporation
Detroit, Michigan

Facts and Methods Pertaining Primarily to Business

HELEN R. JETER
Federal Security Agency
Washington, D. C.

Facts and Methods Pertaining to Sociology, Social Welfare Problems, and Labor Statistics

ERNEST M. FISHER
American Bankers Association
Washington, D. C.

Facts and Methods Pertaining to Financial Institutions

DOUGLAS E. SCATES
Duke University
Durham, North Carolina

Facts and Methods Related to Biometry, Vital Statistics, Psychology, and Education

FREDERICK V. WAUGH
U. S. Department of Agriculture
Washington, D. C.

Facts and Methods Pertaining to Marketing

Secretary-Treasurer

R. L. FUNKHOUSER
210 Normandy Building
1626 K Street, N.W.
Washington, D. C.

Editor

FRANK A. ROSS
210 Normandy Building
1626 K Street, N.W.
Washington, D. C.

Review Editor

GLENN E. MCLAUGHLIN
210 Normandy Building
1626 K Street, N.W.
Washington, D. C.

Associate Editors

F. LESLIE HAYFORD
General Motors Building
New York City

LOWELL J. REED
The Johns Hopkins University
Baltimore, Maryland

GLENN E. MCLAUGHLIN
210 Normandy Building
1626 K Street, N.W.
Washington, D. C.

FREDERICK F. STEPHAN
Cornell University
Ithaca, New York

Biometrics Section

Chairman

C. I. BLISS
The Connecticut Agricultural Experiment Station
New Haven, Connecticut

Secretary

JOSEPH BERKSON
Mayo Clinic
Rochester, Minnesota

CHAPTERS AND DISTRICT REPRESENTATIVES

ALBANY CHAPTER—Edward T. Frankel, *491 State Street, Albany, New York*

ATLANTA CHAPTER—Paul G. Barnett, *Southern Bell Telephone and Telegraph Company, 1501 Hurt Building, Atlanta, Georgia*

AUSTIN CHAPTER—John R. Stockton, *University of Texas, Austin, Texas*

BOSTON CHAPTER—Roswell F. Phelps, *Massachusetts Department of Labor and Industries, Boston, Massachusetts*

CHICAGO CHAPTER—John F. Reinboth, *Commonwealth Edison Company, 72 West Adams Street, Room 1548, Chicago, Illinois*

CINCINNATI CHAPTER—Walter A. Baude, *College of Engineering and Commerce, University of Cincinnati, Cincinnati, Ohio*

CLEVELAND CHAPTER—Edward A. Stephen, *The Ohio Bell Telephone Company, Cleveland, Ohio*

COLUMBUS CHAPTER—Paul J. Daugherty, *Ohio Chamber of Commerce, 1325 Huntington Bank Building, Columbus, Ohio*

CONNECTICUT CHAPTER—Louis M. Nichols, *General Electric Supply Corporation, 1260 Boston Avenue, Bridgeport, Connecticut*

DETROIT DISTRICT CHAPTER

HARRISBURG STATISTICAL SOCIETY—Harry Hoyle, *Pennsylvania Department of Labor and Industry, Harrisburg, Pennsylvania*

LEHIGH VALLEY STATISTICAL SOCIETY

MADISON WISCONSIN CHAPTER

MONTGOMERY CHAPTER—Leonard V. Phelps, *Alabama Department of Health, Montgomery, Alabama*

NEW YORK DISTRICT CHAPTER—Mrs. Helen Slade, *400 East 57th Street, New York City*

PHILADELPHIA CHAPTER—Robert R. Williams, *Department of Research and Statistics, Federal Reserve Bank of Philadelphia, Philadelphia, Pennsylvania*

PITTSBURGH CHAPTER

SAN FRANCISCO CHAPTER—William A. Sturm, *California State Chamber of Commerce, 350 Bush Street, San Francisco, California*

UNIVERSITY OF ILLINOIS CHAPTER—William A. Neiswanger, *University of Illinois, Urbana, Illinois*

WASHINGTON STATISTICAL SOCIETY—Harold D. Kube, *Office of Production Management, New Social Security Building, Washington, D. C.*

A Leading Standard Text

BUSINESS STATISTICS

By JOHN R. RIGGLEMAN
Consulting Economist and Statistician

and IRA N. FRISBEE
Associate Professor of Accounting
University of California at Los Angeles

Second edition. 790 pages, 6 x 9, $4.00

During the past eight years this book has built up an impressive record of adoption in colleges, universities, and business schools throughout the country. Now generally regarded as one of the leading standard texts on the subject, *Business Statistics* stands out prominently in a crowded field. One hundred fifty institutions have adopted the second edition. The following list is representative.

University of Akron
University of Arizona
University of Arkansas
Brooklyn Polytechnic Institute
Bucknell University
Butler University
University of California at Los Angeles
Catholic University of America
Creighton University
Dartmouth College
University of Delaware
De Paul University
Duke University
Harvard University
Loyola University
McGill University
Miami University
University of Nevada
University of Newark
New York University
Northeastern University
Pace Institute
Pennsylvania State College
Rider College
Texas Christian University
Villanova College
University of Virginia
Washington and Lee University
University of Wisconsin
Yale University

Student's Manual for Business Statistics

By IRA N. FRISBEE and WILLARD N. ANDERSON, University of California at Los Angeles. 250 pages, 8½ x 11. $2.00

This helpful book is designed to speed up the problem work, particularly in courses which do not include regular laboratory meetings. The manual provides forms and suggested methods which should reduce materially the time and effort of both students and instructor in connection with mechanical details of set-up, arrangement, and procedure.

Instructor's Manual. *Second edition*

In this manual the chief aim has been to assist the instructor by presenting solutions to mathematical problems and by suggesting points to be considered in the analysis of data in the problems.

Send for copies on approval

McGRAW-HILL BOOK COMPANY, Inc.

330 West 42nd Street **New York, N.Y.**

Please mention the Journal of the AMERICAN STATISTICAL ASSOCIATION *in writing advertisers*

ASSOCIATION MEMBERS!

While attending your Convention, stay at

A HEADQUARTERS HOTEL

Choose One That Is Convenient To Your Meetings

THE BILTMORE

Madison Avenue at 43rd Street

★ ★

THE COMMODORE

Lexington Avenue at 42nd Street

★ ★

THE ROOSEVELT

Madison Avenue at 45th Street

★ ★

These Hotels are directly connected with the Grand Central Terminal and are convenient to all New York's famous attractions. Guest rooms are spacious, luxuriously appointed; cuisine of highest quality; service is skilled and courteous.

ALL ROOMS WITH BATH

SPECIAL CONVENTION RATES

Just Published

STATISTICS FOR SOCIOLOGISTS

By MARGARET JARMAN HAGOOD

Institute for Research in Social Science
University of North Carolina

The purpose of this text is two-fold: to afford an understanding and appreciation of quantitative research methods in order that past and current researches of others in the field may be comprehended and evaluated; and to provide a mastery of the procedures of statistical analysis and especially of the interpretation of the results of statistical analysis.

Invaluable for all Sociologists—Government Employes—Statisticians—Research Workers—Students of the Social Sciences. $4.00

REYNAL & HITCHCOCK, INC.
386 Fourth Ave., New York

THE JOURNAL OF FARM ECONOMICS

Published by The American Farm Economic Association

Editor: T. W. SCHULTZ

IOWA STATE COLLEGE AMES, IOWA

Volume XXIII, November, 1941, Number 4

Some of the major articles are:

Foods for Defense ..... John B. Canning

Fundamental Elements in the Current Agricultural Situation ..... John D. Black

An Agricultural Policy for Hemisphere Defense ..... Francis A. Linville

Social Conflict and Agricultural Programs ..... Kenneth H. Parsons

Training of Personnel for the Rural Social Sciences ..... M. R. Benedict

Economics of Joint Costs in Agriculture ..... S. V. Ciriacy-Wantrup

This Journal, a quarterly, contains in addition, notes reviews of books and articles and a list of recent publications and is published in February, May, August, November by the American Farm Economic Association. Yearly subscription $5.00.

Secretary-Treasurer: **ASHER HOBSON**
Department of Agricultural Economics
University of Wisconsin, Madison, Wisconsin

READY IN DECEMBER

Principles of Punch-Card Machine Operation

How to Operate Punch-Card Tabulating and Alphabetic Accounting Machines

by HARRY PELLE HARTKEMEIER

Professor of Accounting and Statistics, University of Missouri

192 pages, 8½ x 11, $3.00

ONE OF the most significant recent developments in statistics and accounting is the phenomenal increase in the use of punch-card machines. This expansion has been so rapid that its true importance in the educational field is not yet generally recognized.

This book provides all the text and illustrative material necessary to a proper understanding of the fundamentals of the punch-card method. It has been prepared with the cooperation of the International Business Machines Corporation and contains complete wiring diagrams showing every connection necesary for the operation of the four basic models discussed. Although the illustrations are largely from the field of business, this manual is well adapted to an understanding of the applications of tabulating machines to other fields such as business statistics, accounting, economics, education, psychology, agriculture, engineering, home economics and astronomy.

THOMAS Y. CROWELL COMPANY, 432 Fourth Avenue, New York, N.Y.

AMERICAN MARKETING ASSOCIATION

An Association for the Advancement of Science in Marketing

THE American Marketing Association, formed by merging the American Marketing Society with the National Association of Marketing Teachers, began its existence on January 1, 1937, and now has over 1000 active members in all parts of the country.

The principal purposes of the Association are to foster scientific study and research in the field of marketing, to improve the methods and technique of market research and to develop better public understanding and appreciation of marketing problems.

Active Chapters have been established and hold regular meetings in New York, Philadelphia, Washington, Chicago, San Francisco and Los Angeles. Local Chapters are being organized in other cities.

The official publication of the Association, *The Journal of Marketing*, is issued quarterly. Now in its sixth year, the Journal is regarded by marketing practitioners as a source of authoritative marketing information. Many of the contributing authors are leaders in their field of endeavor.

The full membership rate is $5 annually. This includes a subscription to *The Journal of Marketing*. For further particulars, please address the Secretary.

President
Prof. HOWARD T. HOVDE,
University of Pennsylvania,
Philadelphia, Pa.

Treasurer
HOWARD W. GREEN
1001 Huron Road
Cleveland, Ohio

Secretary
Prof. ALBERT HARING
Indiana University
Bloomington, Ind.

THE ECONOMIC ALMANAC for 1941-42

Prepared by the Research Staff of The Conference Board

An authoritative desk book of tested facts about business, labor, government and national resources.

Basic Indicators of Business Activity

Recent Economic Trends: Population, Resources, Industries, National Wealth, National Income, Savings, Consumption, Standards of Living, Private Debt, Public Finance, International Economic Comparisons

Government Organization and Defense Activities: Federal Administrative Agencies—Emergency Defense Organization

Glossary of economic and statistical terms

Comprehensive Index

Expanded and completely revised—464 pages—$4

NATIONAL INDUSTRIAL CONFERENCE BOARD
Dept. S, 247 Park Avenue **New York, N.Y.**

The November 1941 issue of THE ANNALS consists of a series of lectures offered in the summer of 1941 by the University of Michigan Graduate Study Program entitled

PUBLIC POLICY IN A WORLD AT WAR

This volume edited by Dwight L. Dumond, Professor of History at the University of Michigan, presents a coordinated account of the fundamental elements of American domestic and foreign policy, the forces which have shaped the course of international affairs in recent decades, and the relationship of the United States to the present war and prospective peace.

$2.00 per copy ($1.00 to members of the Academy)

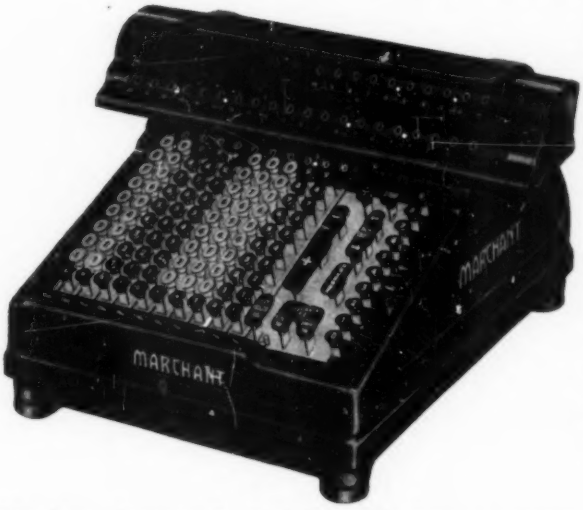
THE AMERICAN ACADEMY OF POLITICAL AND SOCIAL SCIENCE
3457 Walnut Street, Philadelphia, Pa.

Please mention the Journal of the AMERICAN STATISTICAL ASSOCIATION *in writing advertisers*

MARCHANT *"Silent Speed"*

31st Year

CALCULATORS



with the

NEW AUTOMATIC SELECTIVE TABULATION to *any* carriage position

and

COMPLETE CAPACITY CARRY-OVER

Marchant advantages for Statistical Method are too numerous to state in the space of this advertisement. Send for Marchant Methods as below, which contain many new and heretofore unpublished mathematical short-cuts of interest to all statisticians:

MM-45	The Pearson Coefficient with or without Standard Deviations
MM-119	Linear Regression—short-cut equation especially transformed for rapid machine calculation of slope and intercept.
MM-146	Mean and Standard Deviation—data grouped by class intervals.
MM-165	All types of Summations—self-proving, without Charlier Check.
MM-177	Summations of (UX^2) or (X^2) with or without summation (X^2).
MM-178	Short-cut Solution of equations containing $\sqrt{N}$ in numerator or denominator, without need of calculating $\sqrt{N}$.
MM-182	The new Prescott Crout Method of Solving Systems of Linear Equations (with or without symmetry).
MM-110	Continuous Multiplication and Division without manual dial transfers.

Table 56 Square Root by only three calculator entries.
And many others relating to Straight-Line and Curvilinear Interpolation, Differencing, Table Preparation, Cube Root, etc.

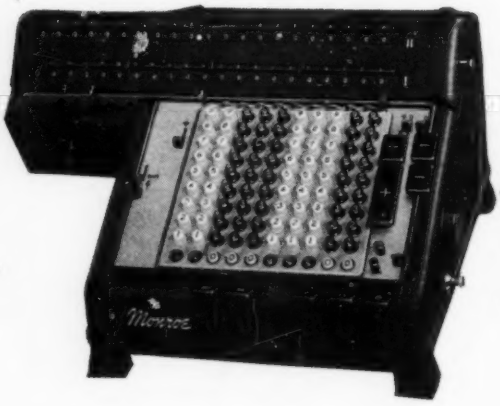
MARCHANT CALCULATING MACHINE COMPANY
HOME OFFICE: OAKLAND, CALIFORNIA, U.S.A.

Sales Agencies and Manufacturers' Service Stations in all Principal Cities

A New Monroe Adding-Calculator

MODEL AA-1

DESIGNED FOR statisticians in the educational governmental research and business statistics fields



DESIGNED BY statisticians based upon over a quarter of a century of service and experience in the use of Monroe Adding Calculators

KNOWLEDGE, based upon statistical facts, is essential today in every line of endeavor. Speed up the flow of vital statistics and insure their accuracy in this era of changing conditions with the assistance of this new Monroe Adding-Calculator. Here are a few statistical accomplishments of this new machine, many of which are exclusive.

Individual X and Y for proof.
Summations X, Y, X^2, Y^2, XY.
Single keyboard for all Set-ups.
Exclusive squaring lock providing single set-up for X and Y.
Automatic carriage positioning for X and Y.
Automatic Negative Multiplication.
Automatic short-cut for X and Y.
Exclusive increased dial capacity provides for unlimited pairs of observations.
Extreme flexibility for all formulae including computation for value of r.

MONROE CALCULATING MACHINE COMPANY, INC., Orange, N. J.

Consult nearest Monroe office for demonstration
or write to Educational Department, Orange, N. J.

FRIDÉN

Automatic Calculators

Accurate Figures in a Hurry

Fridén Automatic Calculators are Sold and Serviced by a Factory trained personnel in over 250 Company controlled Sales Agencies throughout the United States and Canada.

FRIDÉN
CALCULATING MACHINE CO. INC.
FACTORY AND HOME OFFICE – SAN LEANDRO, CALIF

The Annals of Mathematical Statistics

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Edited by S. S. WILKS, A. T. CRAIG, and J. NEYMAN

CONTENTS OF VOLUME XII, NUMBER 3, SEPTEMBER, 1941

	Page
On the Integral Equation of Renewal Theory. Willy Feller	243
On the Joint Distribution of the Medians in Samples from a Multivariate Population. A. M. Mood	268
Samples from Two Bivariate Normal Populations. Chung Tsi Hsu	279
On Randomness in Ordered Sequences. L. C. Young	293
On Certain Likelihood-Ratio Tests Associated with the Exponential Distribution. Edward Paulson	301
On the Mathematically Significant Figures in the Solution of Simultaneous Linear Equations. L. B. Tuckerman	307
On Mechanical Tabulation of Polynomials. J. C. McPherson	317
On the Probability of the Occurrence of at Least m Events Among n Arbitrary Events. Kai Lai Chung	328
Notes:	
A Note on Sheppard's Correction. Cecil C. Craig	339
On the Analysis of Variance in Case of Multiple Classifications with Unequal Class Frequencies. Abraham Wald	346
The Frequency Distribution of a General Matching Problem. T. N. E. Greville	350
On Methods of Solving Normal Equations. Paul G. Hoel	354
Conditions that the Roots of a Polynomial be Less Than Unity in Absolute Value. Paul A. Samuelson	360
Values of Mills' Ratio of Area to Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument. Robert D. Gordon	364

Subscription Rate: $4.00 per annum

Address orders for subscriptions and back numbers to Professor Edwin G. Olds, Secretary, Institute of Mathematical Statistics, Carnegie Institute of Technology, Pittsburgh, Pennsylvania

Copies of the Cumulative Index of Volumes I-X of the *Annals of Mathematical Statistics* can be obtained at $.50 each from the Secretary.

Please mention the Journal of the AMERICAN STATISTICAL ASSOCIATION *in writing advertisers*